A TECHNOLOGY ACCEPTANCE MODEL FOR EMPIRICALLY TESTING NEW END-USER INFORMATION SYSTEMS: THEORY AND RESULTS

by

Fred D. Davis, Jr.
B.S., Industrial Engineering, Wayne State University
(1980)

SUBMITTED TO THE SLOAN SCHOOL OF MANAGEMENT
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF

Ph.D. IN MANAGEMENT

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

December 20, 1985

Signature of Author ______________________________
Sloan School of Management
December 20, 1985

Certified by ______________________________
John C. Henderson, Ph. D
Thesis Supervisor

Accepted by ______________________________
Ph.D. Program Chairman

A TECHNOLOGY ACCEPTANCE MODEL FOR EMPIRICALLY TESTING NEW END-USER INFORMATION SYSTEMS: THEORY AND RESULTS

by

Fred D. Davis, Jr.

Submitted to the Sloan School of Management, M.I.T. on December 20, 1985
in partial fulfillment of the Degree of Ph.D. in Management

ABSTRACT

The goal of this research is to develop and test a theoretical model of the effect of system characteristics on user acceptance of computer-based information systems. The model, referred to as the technology acceptance model (TAM), is being developed with two major objectives in mind. First, it should improve our understanding of user acceptance processes, providing new theoretical insights into the successful design and implementation of information systems. Second, TAM should provide the theoretical basis for a practical "user acceptance testing" methodology that would enable system designers and implementors to evaluate proposed new systems prior to their implementation. Applying the proposed model in user acceptance testing would involve demonstrating system prototypes to potential users and measuring their motivation to use the alternative systems. Such user acceptance testing could provide useful information about the relative likelihood of success of proposed systems early in their development, where such information has greatest value. Based on these objectives, key questions guiding this research include:

(1) What are the major motivational variables that mediate between system characteristics and actual use of computer-based systems by end-users in organizational settings?

(2) How are these variables causally related to one another, to system characteristics, and to user behavior?

(3) How can user motivation be measured prior to organizational implementation in order to evaluate the relative likelihood of user acceptance for proposed new systems?

For user acceptance testing to be viable, the associated model of user motivation must be valid. The present research takes several steps toward establishing a valid motivational model of the user, and aims to provide the foundation for future research that will tend to lead toward this end. Research steps taken in the present thesis include: (1) a fairly general, well-established theoretical model of human behavior from psychology was chosen as a paradigm within which to formulate the proposed technology acceptance model; (2) several adaptations to this paradigm were introduced in order to render it applicable to the present context; (3) published literature in the Management Information Systems and Human Factors fields was reviewed to demonstrate that empirical support exists for various elements of the proposed model, while at the same time the model goes beyond existing theoretical specifications, building upon and integrating previous research in a cumulative manner; (4) measures for the model's psychological variables were developed and pre-tested; (5) a field survey of 100 organizational users was conducted in order to validate the measures of the model's variables, and to test the model's structure, and (6) a laboratory user acceptance experiment of two business graphics systems involving 40 MBA student subjects was performed to further test the proposed model's structure, to test the ability to substitute videotape presentation for hands-on interaction in user acceptance tests, to evaluate the specific graphics systems being tested, and to test several theoretical extensions and refinements to the proposed model.

Thesis Supervisor: John C. Henderson
Title: Associate Professor of Management Science

ACKNOWLEDGEMENTS

I would like to express my appreciation to those who played an important role in my life during the development of this dissertation. To John Henderson, my committee chairman, thank you for holding me to your high standards, imparting to me some of your wisdom about the ways of the academic world, and giving me your friendship. To Tom Malone, thank you for patiently working through many challanging theoretical questions with me and providing much needed inspiration and advice. To John Carroll, thank you for sharing with me your great behavioral research wisdom and ability, and for guiding me away from methodological danger. To Al Silk, thank you for your moral support and encouragement to keep on truckin'. To Dale Goodhue, thank you for your comradeship and willing attention over many an F & T lunch. To Paul Warshaw, thank you for taking me under your wing and teaching me how to do research. To Al Knappenberger, thank you for the early mentorship and for convincing me that I had the ability to get my Ph.D. at M.I.T. And most of all, to my new bride Susan, thank you for the love that made it all worthwhile.

TABLE OF CONTENTS

CHAPTER I INTRODUCTION

The goal of this research is to develop and test a theoretical model of the effect of system characteristics on user acceptance of computer-based information systems. The model, referred to as the technology acceptance model (TAM), is being developed with two major objectives in mind. First, it should improve our understanding of user acceptance processes, providing new theoretical insights into the successful design and implementation of information systems. Second, TAM should provide the theoretical basis for a practical "user acceptance testing" methodology that would enable system designers and implementors to evaluate proposed new systems prior to their implementation. Applying the proposed model in user acceptance testing would involve demonstrating system prototypes to potential users and measuring their motivation to use the alternative systems. Such user acceptance testing could provide useful information about the relative likelihood of success of proposed systems early in their development, where such information has greatest value (Ginzberg, 1981). Based on these objectives, key questions guiding this research include:

(1) What are the major motivational variables that mediate between system characteristics and actual use of computer-based systems by end-users in organizational settings?

(2) How are these variables causally related to one another, to system characteristics, and to user behavior?

(3) How can user motivation be measured prior to organizational implementation in order to evaluate the relative likelihood of user acceptance for proposed new systems?

A longstanding objective of Management Information Systems (MIS) research has been to improve our understanding of the factors that influence successful development and implementation of computer-based systems in organizations (e.g., Keen, 1980). Although there has been no lack of controversy regarding how MIS success should be defined and measured, three variables have consistently emerged as MIS success criteria: actual system usage, user attitudes, and performance impacts (e.g., Alavi & Henderson, 1981; Bailey & Pearson, 1983; Ginzberg, 1981; Ives, Olson & Baroudi, 1983; Lucas, 1975; Swanson, 1974; Zmud, 1979). Considerable MIS research has been devoted to the development of measurement instruments that enable MIS practitioners to assess and monitor MIS success criteria in their organizations (e.g., Bailey & Pearson, 1983; Schultz & Slevin, 1975). Moreover, MIS researchers have sought to provide practitioners with a greater understanding of how they may influence these "success variables" through their control of various policies and decisions, including: (1) choice of system characteristics (e.g., Lucas & Neilson, 1980); (2) choice of development process (e.g., Alavi, 1984); (3) choice of implementation strategy (e.g., Alavi & Henderson, 1981), and (4) nature of support services provided (e.g., Rockart & Flannery, 1983). To a great extent, MIS research is concerned with the development of theories and techniques that permit practitioners to better measure and predict how the decisions under their control affect MIS success. Within this broad context, the present research is concerned with developing techniques for enabling practitioners to assess the impact of one class of managerially controllable variables, system characteristics, on the motivation of members of the intended user community to accept and use new end-user information systems.

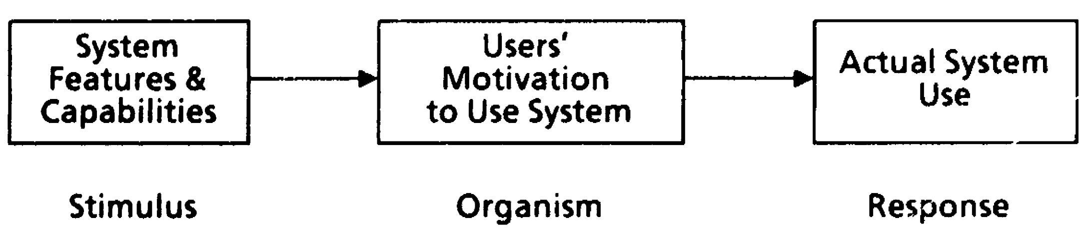
The present research focuses specifically on the class of systems referred to as end-user systems, defined here as systems that are directly used by organizational members at their own discretion to support their work activities. End-user systems represent an increasingly important class of information systems. Direct use of information systems by end-users at all organizational levels has expanded rapidly in recent years and is expected to continue growing strongly into the 1990's (Benjamin, 1982; Rockart & Flannery, 1983; Rockart & Treacy, 1982). The U.S. market for office automation systems was roughly $11.3 billion in 1983, for example, and is expected to grow to $36.6 billion by 1988 (Weizer & Jackson, 1983). The proliferation of personal computing in the early 1980's is indicative of the rapid growth of end-user computing.

Advances in computer technology have clearly been a major force driving this growth, making powerful end-user systems economically attractive. Coupled with this is the growing emphasis that system designers are placing on transforming raw computing capability into systems that fit the needs of end users (Gould & Lewis, 1983; Norman, 1983). In attempting to design more successful systems, developers have found that testing system prototypes with actual prospective users is an effective way of evaluating and refining proposed designs (e.g., Bewley et al., 1983; Card, English & Burr, 1978; Gould, Conti & Hovanyecz, 1983). Although existing protoype testing methodologies are considered valuable (Gould & Lewis, 1985), they suffer from limitations in the techniques used to measure user responses to prototype systems. Namely, current approaches generally do not assess whether users will use the new systems, instead focusing on objective performance criteria (see review in Chapter 3 below) as the basis for choosing among design alternatives. Since end-user systems are primarily used at the discretion of the user (DeSanctis, 1983; Ives, Olson & Baroudi, 1983; Robey, 1979), an important evaluation criterion in many system design situations is whether or not the system under consideration will be used by the target user population. Although actual organizational performance gains are the desired outcome from the use of new information systems, these gains will not be obtained if users fail to adopt the new system. The present research is concerned with the relationship between design characteristics and system use, leaving the use-performance issue aside for the moment. The actual use or non-use of an information system is an important and overlooked issue in the design and selection of information systems.

Conceptual Framework

As discussed above, the proposed model is intended to describe the motivational processes that mediate between system characteristics and user behavior, as depicted in Figure 1.1. System features and capabilities are largely

Figure 1.1 Conceptual Framework



under the control of MIS practitioners: system designers, developers, selectors and managers. Whether a new system is developed in-house for internal use, purchased from an outside vendor, or designed internally for external sale, practitioners have significant influence over the features and capabilities to be included in the target system. The features of the system, in turn, effect the degree to which target users actually use the system. The present research posits that there is an intervening motivational response on the part of the user. Namely, the characteristics of the system effect how motivated users are to use the system, which in turn effects their own actual system use or non-use. Further, the present research aims to develop a model of the motivational variables linking system features with actual use, and to develop measures for these variables. Such measures make it possible to empirically test the proposed model, and may provide the instrumentation needed for applied use of the proposed model in user acceptance testing.

To understand the value of a motivational model of the features-use relationship, consider that information systems practitioners typically need to commit substantial resources when making decisions regarding system characteristics. These resources are at risk to the extent that the target user population may not adopt the new systems. In order to assess the effect of their decisions directly on actual system use, practitioners would need to actually implement the various system alternatives in the organizational setting. In fact, this is frequently done on a limited scale using "test beds" to assess user acceptability of designs (e.g., Johanson & Baker, 1984). Actual organizational implementation has several drawbacks, however. A certain amount of time may need to transpire in order for subjects to begin to integrate the new tools into their work habits (perhaps on the order of several weeks). It is disruptive to the organization, particularly if after the assessment the new systems should be withdrawn from the test users who may have become somewhat dependent on them. Moreover, this approach requires that the test systems be in a much more "finished product" form than is typical for systems at the prototype stage

The present research argues that potential users form motivational tendencies fairly rapidly after being exposed to a new system, and well in advance of the observable behavioral consequences of such tendencies. That is, we hypothesize that demonstrating new systems and their capabilities to potential users in brief testing sessions enables them to form judgments about the applicability of the systems to their jobs, and to form general motivational responses. If true, then measurements of user motivation could be taken from users after a relatively brief exposure to a test system. This would permit practitioners to gather information regarding the comparative acceptability of various alternative systems much earlier in the development process, without the disruptive process of test-bed implementation.

The user acceptance testing process, as envisioned, would consist of: briefly demonstrating a set of alternative new systems to representatives of the intended user population in a laboratory setting (using hands-on interaction and possibly alternative media such as videotapes to demonstrate systems), and measuring their motivation to use the systems in the context of their jobs. Based upon these measurements, the degree of likely acceptance of the system by the users would be predicted.

If user acceptance testing proves successful in explaining user acceptance, it would provide valuable information for system designers and implementors. Designers would be better equipped to evaluate design ideas early in the system development process and make informed choices among alternative approaches. This would enable them to direct development resources toward high priority systems, and reduce the risk of unsuccessful designs. Organizational implementors of systems would be able to systematically involve the users in the process of selecting the systems to be implemented, and identify implementation problems early enough to take corrective actions or cut substantial losses.

Thesis Overview

For user acceptance testing to be viable, the associated motivational model of the user must be valid. The present research takes the first several steps toward the development of a valid motivational model of the user, and establishes a direction for research that will tend to lead toward this end: (1) As a starting point, a fairly general, well-established theoretical model of human behavior from psychology, the Fishbein (1967; Fishbein & Ajzen, 1975) model, was chosen as a foundation upon which to build the technology acceptance model (see Chapter 2, below); (2) several adaptations to the Fishbein model were introduced in order to render it applicable to the present context (see Chapter 2, below); (3) published literature in the Management Information Systems and Human Factors fields was reviewed to demonstrate that empirical support exists for various elements of the proposed model, while at the same time the model goes beyond existing theoretical specifications, building upon and integrating previous research in a cumulative manner (Keen, 1980) (see Chapter 3, below); (4) measures for the model's psychological variables were developed and pre-tested (see Chapter 4, below); (5) a field survey of 100 organizational users was conducted in order to validate the measures of the model's variables, and to test the model's structure (see Chapter 4, below), and (6) a laboratory user acceptance experiment of two business graphics systems involving 40 MBA student subjects was performed to further test the proposed model's structure, to test the ability to substitute videotape presentation for hands-on interaction in user acceptance tests, to evaluate the specific graphics systems being tested, and to test several theoretical extensions and refinements to the proposed model (see Chapter 5, below).

CHAPTER 2. THEORETICAL MODEL

The purpose of this chapter is to specify the proposed technology acceptance model (TAM) and to present an analysis of its theoretical rationale. We begin by presenting an overview of the Fishbein model, which provides most of the theoretical foundation for TAM. Arguments are presented regarding why the Fishbein model provides an appropriate theoretical paradigm in view of the research objectives. Next, TAM is specified, followed by a detailed discussion of theoretical issues that were considered in its development, including its relationship to the Fishbein model.

Fishbein Model Overview

The Fishbein (1967) model was chosen as the reference paradigm within which the proposed technology acceptance model is developed. The model, originally specified by Fishbein (1967) and extensively analyzed and refined by Fishbein and Ajzen (1975), is defined using three equations. The first equation indicates that an individual's intention to perform a given behavior (BI_{act}) is the immediate causal determinant of his or her overt performance of that behavior (B), and that an individual's intention is jointly determined by his or her attitude toward performing the behavior (A_{act}) as well as the perceived social influence of people who are important to the individual (SN_{act}):

(1) $$B \approx BI_{act} = w_1 A_{act} + w_2 SN_{act}$$

where

B = behavioral criterion

BI_{act} = behavioral intention regarding behavior B

A_{act} = attitude toward behavior B

SN_{act} = subjective norm regarding behavior B

w_1, w_2 = importance weights

Behavioral intention (BI) has typically been defined as an individual's subjective probability that he or she will perform a specified behavior (e.g., Fishbein & Ajzen, 1975, p. 288). Attitude refers to an individual's degree of evaluative affect toward the target behavior (e.g., Fishbein & Ajzen, 1975, p. 216). Subjective norm refers to "the person's perception that most people who are important to him think he should or should not perform the behavior in question" (Fishbein & Ajzen, 1975, p. 302). The importance weights are estimated via multiple regression to reflect the relative causal influence of the attitudinal and normative components in a given situation, and are expected to vary across situations.

The second equation implies that an individual's attitude toward a given behavior is a function of the perceived consequences of performing the behavior multiplied by the evaluations of those consequences:

(2) $$A_{act} = \Sigma_{i=1,n}\ b_i e_i$$

where

b_i = belief that performing behavior B will result in consequence i

e_i = evaluation of consequence i

n = number of salient beliefs

This equation is based on the expectancy-value model of attitude posited by Fishbein (1963) which was built upon the earlier work of Rosenberg (1956). Beliefs are defined as the person's subjective probability that performing the target behavior will result in salient consequence i. The evaluation terms refer to "an implicit evaluative response" to the consequence (Fishbein & Ajzen, 1975, p. 29). Fishbein & Ajzen (1975, p. 216) theorize a tight relationship between beliefs and attitudes: " In our conceptual framework, as a person forms beliefs about an object, he automatically and simultaneously acquires an attitude toward that object.". Equation 2 represents an information-processing view of attitude formation and change, which argues that attitudes are altered only through changes in the individual's belief structure (e.g., Ajzen and Fishbein, p. 253).

The third equation specifies that the individual's subjective norm is a function of "the perceived expectations of specific referent individuals or groups, and by the person's motivation to comply with those expectations" (Fishbein & Ajzen, 1975, p. 302):

(3) $SN_{act} = \sum_{j=1,m} nb_j\, mc_j$

where

nb_j = Normative belief that referent j wants subject to perform behavior B

mc_j = Motivation to comply with referent j

m = Number of salient referents

This is the least-understood part of the model, since "Very little research...has dealt with the formation of normative beliefs"(Fishbein & Ajzen, 1975, p. 304). It is frequently argued that normative beliefs may be incorporated under the expectancy-value attitude component, insofar as the subject may regard complying with the wishes of an important referent as a salient consequence of the target behavior. While agreeing that some normative beliefs may indeed come under the attitudinal component, Fishbein & Ajzen (1975, p. 304) claim that "it is useful to maintain the distinction between beliefs about the consequences of performing a behavior and beliefs about expectations of relevant referents." In addition, empirical studies of the model have shown that the normative component often has a significant effect on intention along with attitude (Ajzen & Fishbein, 1980).

The Fishbein model does not specify which beliefs (i.e., perceived consequences, see equation 2 above) are operative for a given context. Instead, researchers using the model must first identify the beliefs to be included for the situation they will be addressing. Fishbein & Ajzen (1975, p. 218) point out that "Although a person may hold a relatively large number of beliefs about a given object, it appears that only a relatively small number of beliefs serve as determinants of his attitude at any given moment." Those beliefs that exert influence on one's attitude are referred to as salient beliefs. Fishbein & Ajzen (1975, p. 218) suggest eliciting salient beliefs using a free response approach wherein subjects are asked to list the consequences of performing the target behavior that come to mind. At the same time, they point out: "It is possible, however, that only the first two or three beliefs are salient for a given individual and that individual beliefs elicited beyond this point are not primary determinants of his attitude (i.e., are not salient). Unfortunately, it is impossible to determine the point at which a person starts eliciting non-salient beliefs (Fishbein & Ajzen, 1975, p. 218)." As a rule of thumb, Fishbein & Ajzen (1975, p. 218) suggest that five to nine beliefs be elicited, since: "Research on attention span, apprehension, and information processing suggests that an individual is capable of attending to or processing only five to nine items of information at a time" (e.g., G.A. Miller, 1956; Woodworth and Schlosberg, 1954; Mandler, 1967). Since the set of salient beliefs is expected vary across individuals, they suggest using the modal salient beliefs for the population, obtained by taking the beliefs most frequently elicited from a representative sample of the population. A similar elicitation procedure is recommended for the elicitation of salient referents.

An important characteristic of the Fishbein paradigm is the argument that in order to obtain a correct specification of the causal determinants of behavior, the psychological variables of the model should be defined and measured at a level of specificity that corresponds to the behavioral criterion to be explained. That is, the variables of the model should be worded in a way that is parallel to the target behavior in terms of target, action, context, and time frame elements (Fishbein & Ajzen, 1975, p. 369; Ajzen & Fishbein, 1980, p. 34). For example, Ajzen and Fishbein (1980, p. 43) point out: "Imagine that we want to predict, for each respondent in a sample, whether he or she will buy a color television set. Further, suppose we decide to wait a year before measuring whether the behavior has occurred. It can be seen that this criterion specifies an action (buying), a target class (color television sets), and a time period (the year in question), but it leaves the contextual element unspecified. The only measure of intention that corresponds exactly to this behavioral criterion is a measure of the person's intention 'to buy a color television set within a year.' If we had decided to return in six months to record the behavior, the corresponding intention would be the intention 'to buy a color television set within the next six months.'" Similarly, Fishbein and Ajzen (1975) argue that the relationships between beliefs, evaluations, attitudes, subjective norm, normative beliefs and motivations to comply specified by the model will only obtain if these elements correspond in specificity with the behavioral criterion.

A brief discussion specifically about an individual's "attitude toward the object" is warranted. If we are using the Fishbein framework to model the determinants of a person's behavior with respect to a target object, then the appropriate attitude to measure is the person's attitude toward performing the behavior with respect to the object (A_{act}), and not their attitude toward the object per se (A_o). A_{act} corresponds in specificity to the behavioral criterion in terms of the action element, whereas A_o does not (one should similarly ensure correspondence in context and time elements as well). Fishbein and Ajzen (1974) demonstrated that although A_o is strongly linked to general patterns of behavior relative to the attitude object, it is much less able than A_{act} to predict specific behavioral criteria involving the object. The expectancy-value attitude models of Peak (1955), Rosenberg (1956) and Fishbein (1963) were object-based and not behavior-based attitude models. Hence, the belief structures being dealt with pertained to the perceived attributes of the object as opposed to anticipated behavioral consequences (e.g., Fishbein, 1963; Fishbein & Ajzen, 1975). Since a great deal of prior attitude research was concerned with the measurement of A_o and perceived attributes of attitude objects, it was frequently unable to identify clear linkages between the measured "attitudes" and specific behavior (for review, see Wicker, 1969). One of the major contributions of the Fishbein paradigm was to resolve inconsistent findings where such lack of correspondence was at fault (e.g., Ajzen & Fishbein, 1977). The Fishbein model views A_o as an external variable, exerting influence over intention only through its effect on beliefs about the behaviors consequences, evaluations of the consequences, normative beliefs, motivations to comply and importance weights, as discussed above (Fishbein & Ajzen, 1975, p. 315-316; Ajzen & Fishbein, 1980, p. 84).

To the extent that the measured intention fails to correspond in specificity to the target behavior, its causal relationship to the behavior is attenuated, resulting in reduced predictive capability. In addition to correspondence in levels of specificity, Fishbein and Ajzen (1975, p. 370-71) discuss two other conditions under which the ability of the intention variable to predict behavior will be reduced. First, as the time between the measurement of a person's intention and the observation of their behavior increases, the likelihood that their intention may change is increased, reducing the overall predictiveness of the original intention. Second, to the extent that the behavioral criterion is not under the actor's volitional control, their reduced ability to carry out their intention translates into reduced behavioral predictiveness. Lack of volitional control may arise in cases where the individual lacks the ability or resources to carry out an intended behavior.

The Fishbein model asserts that external variables, such as the characteristics of the behavioral target, influence behavioral intentions only indirectly by influencing the individual's beliefs, evaluations, normative beliefs, motivation to comply, or the importance weights on the attitudinal and subjective norm components (Fishbein & Ajzen, p. 307). External variables encompass all variables not explicitly represented in the model, and include demographic or personality characteristics of the actor, the nature of the particular behavior under consideration, characteristics of referents, and prior behavior, and persuasive communication. Various behavioral change strategies discussed by Fishbein and Ajzen (1975, p. 387-509) are founded upon the principle that an individual's intention may be influenced primarily by influencing his or her beliefs.

There are several attractive characteristics of the Fishbein model as a paradigm for the present research. A major advantage often cited in favor of the Fishbein model is that it integrates a number of previously disjoint theories concerning the relationships between beliefs, attitudes, intentions, and behavior. Fishbein's model is an adaptation of Dulaney's (1961) theory of propositional control, which was developed in the context of laboratory experiments on verbal conditioning and concept attainment. Fishbein and Ajzen (1975) present a detailed analysis of the Fishbein model in relation to major existing theories arguing that the Fishbein formulation is quite similar to alternative attitude models from such perspectives as: learning theory (e.g., Doob, 1947; Staats & Staats, 1958); expectancy-value theory (e.g., Atkinson, 1957; Edwards, 1954; Rotter, 1954, Tolman, 1932); consistency theory (e.g., Festinger, 1957; Heider, 1946; Rosenberg, 1960), and attribution theory (e.g., Heider, 1958; Jones & Davis, 1965; Kelley, 1967). The Fishbein model is similar in structure to other major motivation theories as well (e.g., see Vroom, 1964; Weiner, 1985). In addition, the Fishbein model is very explicit regarding the definitions of, operationalizations of, and causal relationships among the variables being addressed compared to many alternative theoretical perspectives. A substantial body of empirical results has accumulated which generally provides support for the model specification (Ajzen & Fishbein, 1980; Fishbein & Ajzen, 1975; Ryan & Bonfield, 1975). The Fishbein model has been widely used in applied research spanning a variety of subject areas (Brinberg & Durand, 1983; Davidson & Morrison, 1983; Hom, Katerberg & Hulin, 1979; Jaccard & Davidson, 1972; Manstead, Proffitt & Smart, 1983), while at the same time stimulating a great deal of theoretical research aimed at understanding the model's limitations, testing key assumptions, and analyzing various refinements and extensions (Bagozzi, 1981, 1982, 1984; Bentler & Speckart, 1979; Ryan, 1982; Saltzer, 1981; Warshaw, 1980a, 1980b; Warshaw & Davis, 1984, 1985, in press; Warshaw, Sheppard & Hartwick, in press).

The Fishbein model appears well-suited to the present research objectives. It provides a well-founded theory of the motivational linkages between external stimuli, of which system characteristics are an instance, and resulting behavior.

Moreover, the model provides criteria for developing operational measures for observing these motivational phenomena prior to their behavioral manifestation.

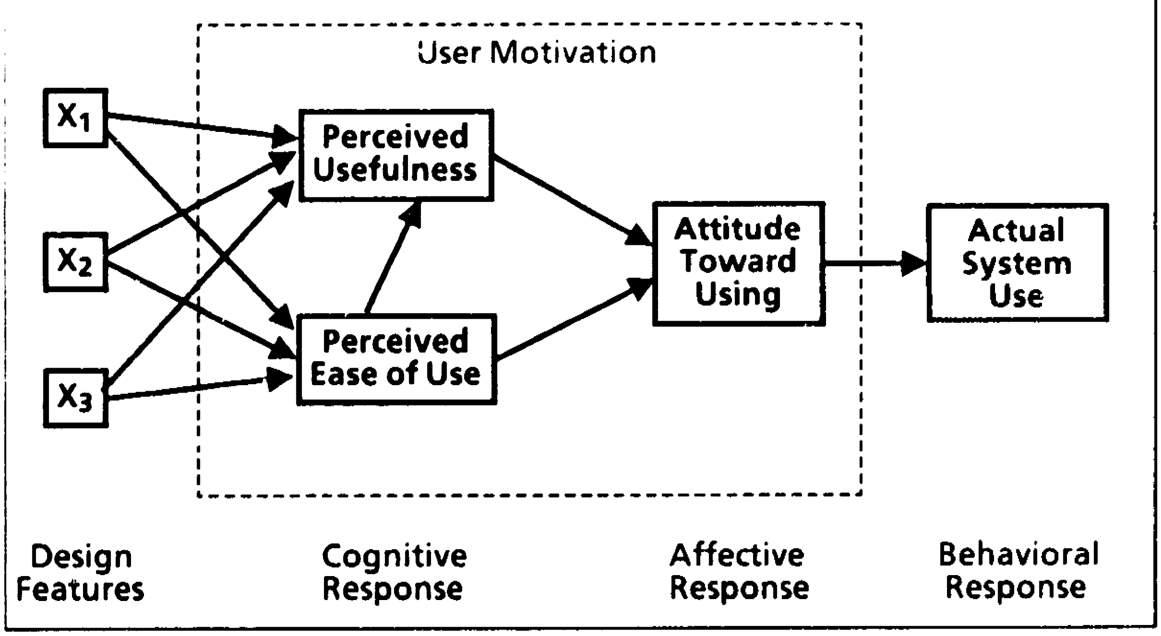
A broader advantage of the Fishbein model is that it is capable of integrating numerous theoretical perspectives from psychology which have previously been employed in MIS acceptance research. In addition, using the Fishbein model in MIS provides the opportunity to take advantage of new theoretical developments and extensions in the reference discipline as they become available.

Technology Acceptance Model

Model Specification

The proposed technology acceptance model is shown in Figure 1, with arrows representing causal relationships. Alternative systems are represented using a set of binary "design feature" variables. According to the model, a potential user's overall attitude toward using a given system is hypothesized to

Figure 1. Technology Acceptance Model



be a major determinant of whether or not he actually uses it. Attitude toward using, in turn, is a function of two major beliefs: perceived usefulness and perceived ease of use. Perceived ease of use has a causal effect on perceived usefulness. Design features directly influence perceived usefulness and perceived ease of use. Since design features fall into the category of external variables within the Fishbein paradigm (as discussed previously), they are not theorized to have any direct effect on attitude or behavior, instead affecting these variables only indirectly through perceived usefulness and perceived ease of use. Following the Fishbein model, the relationships of the model are theorized to be linear. The model can be expressed using the following four equations:

(1) $EOU = \sum_{i=1,n} \beta_i X_i + \varepsilon$

(2) $USEF = \sum_{i=1,n} \beta_i X_i + \beta_{n+1} EOU + \varepsilon$

(3) $ATT = \beta_1 EOU + \beta_2 USEF + \varepsilon$

(4) $USE = \beta_1 ATT + \varepsilon$

where

X_i = design feature i, i = 1,n
EOU = perceived ease of use
USEF = perceived usefulness
ATT = attitude toward using
USE = actual use of the system
β_i = standardized partial regression coefficient
ε = random error term

Use refers to an individual's actual direct usage of the given system in the context of his or her job. Thus, use is a repeated, multiple-act behavioral criterion (e.g., Fishbein & Ajzen, 1975, p. 353) that is specific with respect to target (specified system), action (actual direct usage) and context (in person's job), and non-specific with respect to time frame. Attitude refers to the degree of evaluative affect (e.g., Fishbein & Ajzen, 1975, p. 216) that an individual associates with using the target system in his or her job. Therefore, the definition and measurement of attitude corresponds in specificity with the definition of the behavioral criterion, as recommended by Ajzen and Fishbein (1977). Perceived usefulness is defined as "the degree to which an individual believes that using a particular system would enhance his or her job performance." Perceived ease of use is defined as "the degree to which an individual believes that using a particular system would be free of physical and mental effort."

Perceived ease of use is hypothesized to have a significant direct effect on perceived usefulness, since, all else being equal, a system which is easier to use will result in increased job performance (i.e., greater usefulness) for the user. Given that a non-trivial fraction of a user's total job content is devoted to physically using the system per se, if the user becomes more productive in that fraction of his or her job via greater ease of use, then he or she becomes more productive overall. Thus, characteristics of the system may indirectly influence usefulness by affecting ease of use.

Theoretical Rationale for Technology Acceptance Model

The purpose of this section is to discuss a variety of theoretical considerations upon which the proposed model is founded. Particular attention is directed to the relationship between the technology acceptance model and the Fishbein model, the latter of which provides a major conceptual foundation for the present research. The technology acceptance model as specified above departs in several ways from the standard Fishbein model. The manner in which beliefs (perceived consequences) are specified, modeled and measured differs from the recommended Fishbein approach. In addition, the subjective norm and behavioral intention variables, although central elements of the Fishbein paradigm, are omitted from the present model. The nature of, and rationale for, these adaptations are discussed.

Belief summation. Recall that in the Fishbein paradigm the relationship between beliefs and attitude is modeled using the following equation:

$$A_{act} = \sum_{i=1,n} b_i e_i$$

The relationship between beliefs and attitudes is generally assessed by computing the summation on the right-hand side of the above equation and computing its correlation with A_{act}. This is equivalent to computing the standardized regression coefficient of the following regression equation (Pindyck & Rubenfeld, 1981):

$$A_{act} = \alpha + \beta\left[\sum_{i=1,n} b_i e_i\right] + \varepsilon$$

In this approach, the summed belief-evaluation term is conceptually treated as a single independent variable, and the regression (or correlation) coefficient is interpreted as the overall effect of beliefs on attitude.

In contrast, the present model represents each belief separately in the regression equation for A_{act} (see equation 3 above). Modeling the belief structure in a disaggregated way using multiple regression enables one to compare the relative influence of different beliefs in determining attitude toward using. Regression has long been used in this manner within a variety of theoretical domains to model the processes that subjects use for integrating distinct information elements when forming evaluative judgments (for review, see Slovic & Lichtenstein, 1971). Regression-based models appear to be able to accurately model the importance of cues used in judgmental processes observed via fine-grained protocol analyses, even when highly non-compensatory judgmental processed are employed by subjects (Einhorn, Kleinmuntz & Kleinmutz, 1979; see also Johnson & Meyer, 1984). Although Fishbein and Ajzen (1975, p. 158-150) acknowledge the virtues of multiple regression for modeling the effects of beliefs on other beliefs, they do not use it for analyzing the relationship between beliefs and attitude.

In discussing the relative importance of the individual belief items i, Fishbein and Ajzen (1975, p. 241) admit they: "have essentially assumed that the weight is 1.0 and can thus be neglected." They argue that the size of the multiple correlation between beliefs and attitude does not generally improve when importance weights are added. However, due to the robust characteristics of linear models, under conditions that obtain in most expectancy-value contexts (i.e., correlated independent variables monotonically related to the dependent variable), unit weights are capable of providing accurate predictions despite substantial deviations from the true regression parameters (Dawes & Corrigan, 1974; Einhorn & Hogarth, 1975; Wainer, 1976). Therefore, it is erroneous to conclude that unit weights represent a good approximation of the actual structural parameters based on the observation that the explained variance is approximately the same. Since we are concerned with accurately estimating the effects of perceived usefulness and perceived ease of use per se, in addition to the multiple correlation for the A_{act} equation, it is more appropriate to estimate regression coefficients for each belief separtately rather than to assign them unit weights. Although we do not expect the overall proportion of explained variance to significantially surpass that of a unit-weighted model, the estimated regression weights are an important source of diagnostic explanatory information which enables the researcher to gauge the relative influence of perceived usefulness and ease of use in determining attitudes and behavior.

In addition to gaining information about about the relative influence of ease of use and usefulness on attitude, representing beliefs separately in the A_{act} equation (equation 3) is consistent with the fact that ease of use and usefulness are treated as separate dependent variables in equations 1 and 2, respectively. By representing beliefs separately in equations 1 and 2, we are able to assess the effect of system characteristics on each belief, apart from one another, and to assess the influence of ease of use on usefulness. The summed belief structure of the Fishbein model does not provide an appropriate basis for assessing the impact of design features on individual beliefs, and is likely to substantially distort the actual underlying effect. For example, in cases where a design feature increases usefulness while at the same time decreasing ease of use, the countervailing perceptual effects may cancel each other out, resulting in an incorrect no difference conclusion. In cases where significant effects of design features on the summed belief structure are observed, it is impossible to attribute the effect to individual beliefs or specific combinations of beliefs making up the summation.

Some researchers have used "self-reported" importance weights in place of estimated regression coefficients. Marketing researchers in particular tend to use self-reported importance weights in multiattribute attitude models (e.g., see reviews by Wilkie & Pessemier, 1973; Ryan & Bonfield, 1975; and Shocker & Srinivasan, 1979). However, based on a review of several marketing studies, Bass & Wilkie (1973) conclude that self-reported importance weights seldom add significant explanatory power, and frequently degrade prediction by a substantial margin. Although the issue remains controversial, a potential explanation for this phenomenon is the frequently expressed view that people appear to lack self-insight regarding the importance they actually attach to various cues in forming judgements (Einhorn & Hogarth, 1981; Nisbett & Wilson, 1977; Slovic & Lichtenstein, 1971). It should be pointed out that researchers generally distinguish self-reported importance weights from the belief evaluation term (e_i) of the Fishbein model (e.g., Fishbein & Ajzen, 1975, p. 228; Ryan & Bonfield, 1975, p. 120). Thus, the use of self-reported importance weights does not appear warranted.

In sum, representing beliefs separately in equations 1-3 provides greater diagnostic and explanatory information regarding the process by which systems feature affect user behavior than would be possible if beliefs were handled in aggregate, as typically done in applications of the Fishbein model. Moreover, the estimation of the importance weights for ease of use and usefulness in equation 3 using regression has substantial advantage relative to using either unit or self-reported weights.

Belief evaluation term. Another key difference between the expectancy-value attitude model posited by Fishbein and the attitude equation being employed above (equation 3) is that, whereas in the Fishbein model each belief is multiplied by its corresponding evaluation term, no evaluation term is employed in the present model. Researchers have argued that to employ such a multiplication term in a regression or correlation analysis assumes that the multiplied terms are scaled at the ratio level of measurement (Bagozzi, 1984; Ryan & Bonfield, 1975; Schmidt, 1973). Since measurements of the type used in operationalizing the Fishbein model constructs are generally interval scaled, this assumption is unteneble, and allowable linear transformations will "change the relationship of the product term with the criterion... For example, a positive correlation can be made to be zero or negative merely through the addition of a constant to measurements of one of the variables in the product term (Bagozzi, 1984, p. 296)." Since there is no fixed zero point in an interval measure, this renders the magnitude of the product term uninterpretable.

To circumvent this problem, Bagozzi (1984) estimated the effect of the product of beliefs and evaluations on attitude using hierarchical regression where the main effects of each term of the product were entered into the regression along with the product term itself. Using this approach, the estimated coefficient of the product term is invariant under linear transformations of beliefs and evaluations. When Bagozzi (1981) applied this analysis to his blood donation data (e.g., Bagozzi, 1982), he found no significant effect for the interaction term, concluding that "There is no evidence, then, that the sum of beliefs times evaluations combine multiplicatively in this case." Fishbein and Ajzen (1975, p. 227) contend that several studies have shown "attitudes can be estimated more accurately by considering both belief strength and evaluation of associated attributes (i.e., from $\Sigma\, b_ie_i$) than by using the sum of the beliefs ($\Sigma\, b_i$) or the sum of the evaluations ($\Sigma\, e_i$)...One exception occurs when the e's are either all positive or all negative. In this case, $\Sigma\, b_i$ alone will tend to be highly correlated with the attitude." However, any set of salient beliefs may be transformed to "all positive" beliefs by reversing the data coding scheme for negatively valenced beliefs. Thus, Fishbein and Ajzen's observation that multiplying beliefs by corresponding evaluations only improves predictiveness when the belief set contains a mixture of positively and negatively valenced beliefs suggests that the evaluation terms may simply serve as a mechanism for reversing the coding of negative beliefs, and that reversing the data coding scheme in the first place may obviate the need for the evaluation term. Nevertheless, in both Bagozzi's (1982) blood donation model and the present context, the beliefs being modeled are homogeneous with respect to evaluative orientation, i.e., all of Bagozzi's beliefs were negatively valanced, whereas both perceived usefulness and perceived ease of use are positively valenced. Hence we would not expect a moderating role for the evaluation term even according to Fishbein and Ajzen's logic. Given the disadvantages introduced by the failure of the belief evaluation terms to possess the ratio scale property needed for employing them multiplicatively with beliefs, coupled with the apparent lack of advantage in terms of improved explanation, it was decided to omit them from the present model.

Relationships between beliefs. The present model posits a causal relationship between perceived ease of use and perceived usefulness. However, the Fishbein model does not explicitly represent relationships between beliefs; instead, as discussed previously, beliefs (times evaluations) are given equal unit weights and summed together, without regard to relationships that may exist among them. Ironically, in their theorizing about the processes of belief formation, Fishbein and Ajzen (1975) place great emphasis on relationships between beliefs. Fishbein and Ajzen draw the distinction between "descriptive beliefs"which are formed based on directly observable objects or events and "inferential beliefs" that go beyond directly observable phenomena. From this perspective, perceived ease of use may be seen as largely a descriptive belief in the context of a user acceptance test, formed based on the subjects' direct experience with the target system. Some inferential processes may influence ease of use perceptions since subjects may have to speculate beyond the short tutorial exposure to predict their ultimate mastery of the target system, taking into consideration their own abilities and past experiences. In contrast, perceived usefulness is considered much more inferential in nature, requiring subjects to estimate the effect of the system on their job performance in the absence of any direct experience of using the system in their job.

Fishbein and Ajzen (1975, Ch. 5) discuss a wide variety of theoretical models regarding the formation of inferential beliefs that address relationships among beliefs, including the effects of descriptive beliefs on inferential beliefs. They cover several theories, including: impression formation (Asch, 1946), cue utilization (Slovic & Lichtenstein, 1971), information integration (Anderson, 1970), multiple cue learning (Brunswik, 1955; Hammond & Summers, 1972), attribution (Heider,1958; Kelley, 1973), and the subjective probability model of Wyer and Goldberg (1970). Thus, the modeling of relationships between beliefs does not appear to be inconsitent with the major theories of belief formation and change upon which the Fishbein model is founded.

Belief salience. As discussed previously, Fishbein and Ajzen recommend using a qualitative free-response elicitation procedure to identify the salient beliefs of a subject population with respect to a given behavior by asking subjects to "list the characteristics, qualities, and attributes of the object or the consequences of performing the behavior" (p. 218). One possible criticism of the present approach is that there is no guarantee that the specified beliefs are in fact salient, since the recommended elicitation procedure was not employed. Two observations tend to mitigate this concern. First, the notion that the recommended elicitation procedure truly identifies the salient beliefs (i.e., those that are influential in attitude formation) is an assumption which has received little validation. One study addressing the issue was reported by Jaccard and Sheng (1985), who applied the Fishbein approach and indexed the importance of a given consequence for an individual according to the order in which it was elicited from the individual. The computed correlations between this elicitation importance index and standardized regression coefficients were found to be -.071 and -.429 for career and birth control decisions, respectively. Jaccard and Sheng's analysis raises questions about the validity of the Fishbein elicitation procedure, and implies that the beliefs resulting therefrom should not automatically be assumed to be the one's most influential in determining the individual's behavioral decision.

The second observation is that the beliefs specified a priori in the present model are based on considerable previously published theoretical and empirical articles that span a wide range of system types and user populations. Perceived usefulness and perceived ease of use have been repeatedly identified as important issues governing user acceptance processes (see Chapter 3). Accordingly, they appear quite appropriate given our objective to develop a general model applicable across many contexts. In constrast, the Fishbein elicitation approach would require eliciting statements of belief from specific subjects when they are asked to think about using a specific system (or set of alternative systems). The set of beliefs that would result from such an elicitation would likely contain some belief items that are idiosyncratic to the subject population, target system or usage context relative to which the beliefs were elicited. Moreover, it would not be possible to separate general beliefs from those that are context-specific. Fortunately, in the present research situation, the existing theoretical and empirical literature provides an important source of information about the beliefs that are expected to be salient in general. Since these studies span a wide range of user populations, systems and usage contexts, the risk of identifying idiosyncratic beliefs is reduced, and the probability that the beliefs which are being tapped are salient is increased.

Fishbein and Ajzen argue that there are usually five to nine salient beliefs for an individual in a given situation. However, researchers have begun to reject the view that each elicited belief corresponds to a distinct belief construct, observing that individual items frequently correlate highly with one another (e.g., Bagozzi, 1982; Hauser & Simmie, 1981; Hauser & Urban, 1977; Holbrook, 1981). Factor analysis is often used to identify the underlying belief dimensions (constructs), although multidimensional scaling is also applied (Shocker & Srinivasan, 1979; Silk, 1969). It has generally been found that a relatively small number of such belief dimensions are operative in a given situation (generally 2-4, e.g., Bagozzi, 1982; Silk, 1969). The conceptual definition of the underlying construct is usually inferred from the content of items loading on the dimension, and the items are frequently treated as measures of the dimension.

The qualitative elicitation and factor analysis of belief items is a valuable technique for identifying belief constructs for multiattribute models, and is especially valuable in circumstances wherein the researcher does not have an a priori model of the operative belief constructs (Bagozzi, 1983). The technique loses some of its appeal in cases where specific belief constructs are specified a priori based on theoretical considerations, such as in the present model. Given the existence of a set of a priori beliefs to be modeled, the use of such elicitation/factor analysis is not needed for belief identification. For measure development, the a priori specification of belief constructs permits the researcher to develop valid and reliable measures specifically tailored for those constructs (Cook & Campbell, 1979; Nunnally, 1978).

The possibility remains, of course, that perceived ease of use and perceived usefulness do not represent a complete specification of the beliefs which are salient in a given situation. Rather than assume that these perceptions represent a complete salient set, we view this as a hypothesis to be tested. Perceived usefulness and perceived ease of use will be regarded as salient to the extent that they exert a causal influence on attitude toward using. In addition, failure of these two beliefs to fully mediate between system characteristics and attitude toward using may suggest that a salient belief has been omitted.

Belief Measurement. Multi-item meaurement scales for perceived usefulness and perceived ease of use will be developed and validated according to recommended procedures from the psychometric literature (e.g., Bohrnstedt, 1970; Nunnally, 1978). The conceptual definitions for usefulness and ease of use will serve as the basis for generating an initial pool of measurement items for each construct based on prior literature. The wording of the items will be pre-tested to verify their correspondence with the underlying conceptual variables they are intended to measure. The items will be operationalized using scale formats recommended by Ajzen & Fishbein (1980, Appendix A). Finally, a survey will be performed to verify the reliability and validity of the scales. This procedure, discussed in greater detail in Section 4, has a number of advantages relative to the belief measurement procedure suggested by Fishbein & Ajzen (1975; see also Ajzen & Fishbein, 1980). Using the recommended procedure, elicited beliefs are directly converted to belief measures using the recommended standard scale formats (e.g., Ajzen & Fishbein, 1980). This has the problem that there is only a single item for measuring each belief. Unfortunately, single-item measurement scales are generally not reliable and valid (Cook & Campbell, 1979; Nunnally, 1978). In contrast, the present approach places emphasis on the development of valid, reliable scales.

Subjective norm. The subjective (social) norm component of the Fishbein model is not included in the specification of the technology acceptance model since, in the applied user acceptance testing context for which the proposed model is being developed, no information will be available to subjects pertaining to the expectations of their salient referents regarding their usage of the target system. Fishbein and Ajzen (1975, p. 304) theorize that normative beliefs can be formed in two ways: "First, a given referent or some other individual may tell the person what the referent thinks he should do, and the person may or may not accept this information. Second, the person may observe some event or receive some information that allows him to make an inference about a given referent's expectations." In a user acceptance test, subjects will typically be seeing the target systems (generally new system prototypes) for the first time, and will therefore not have been able to receive cues from referents upon which to draw normative inferences. This implies that no relevant perceived social normative influences would exist at the time of user acceptance testing.

When questioned about social normative influences in the absence of such influences, subjects may either correctly indicate that they do not have a normative belief either way, or attempt to guess what the social normative influences of their salient referents would be. The former should result in the subjective norm having no influence on intentions or behavior, whereas the latter could introduce error and ambiguity into the measurement of subjective norm. Although such guessing falls outside of the realm of normative belief formation processes theorized by Fishbein and Ajzen (1975), it is interesting to speculate about how subjects may go about guessing what the expectations of their salient referents would be regarding their use of the target systems. Subjects may simply project their own attitudes upon their salient referents, in which case their attitudinal and normative components should be roughly equivalent, making it unlikely that the social normative component will add explanatory power above and beyond attitude alone. Alternatively, subjects may base their guesses upon social normative cues that have been received during actual prior social interaction with salient referents pertaining not to the target system, but to other systems or to the category of which the target system is a member (e.g., "my boss wants me to use graphics more, and this is a graphics system, therefore, he is likely to want me to use this system"). Such "anchoring" norms may potentially yield some useful insights, although they do not fully comply with the correspondence criteria of the Fishbein model which suggest that the social normative construct should pertain to the specific target system being addressed. The extent to which normative judgements made in the absence of specific psychosocial cues are influenced by subject attitude, anchoring norms, and error is not presently known. This limited theoretical understanding suggests that attempting to model such judgements at the present time would be premature.

Potential extraneous sources of normative influence that may be operative in the laboratory setting are subjects' perceived expectancies of the experimenter (e.g., Cook & Campbell, 1979, p. 67), and social influences of other participants. Such influences represent undesired experimental artifacts which are not representative of the social influences that would occur naturally in the organizational setting. Coincidintally, it was precisely such experimenter expectancies that Dulaney (1961) was modeling with the subjective norm variable in his propositional control model, which became the forerunner of the Fishbein model. In order to minimize such nuisance factors, precautions should be taken to reduce experimenter expectancies and social interaction during user acceptance testing experiments.

Behavior intention. The behavioral intention (BI) variable is omitted from the model as well. The main reason for this is that intention reflects a decision that the person has made, and as such gets formed through a process of mental deliberation, conflict and commitment that may span a significant time period (Einhorn & Hogarth, 1981; Janis & Mann, 1977; Warshaw & Davis, 1985). In general, this time period is expected to be proportional to the importance of the decision. The decision whether or not to become a user of a new information system in one's job would generally be regarded as a fairly important decision. In the user acceptance testing context, measurements of subjects' motivation to use a new system would take place directly after demonstrating the system to the user. Thus, the time required to form an intention would not be expected to elapse prior to measurement.

Intention is generally a better predictor than attitude when an intention has been formed (i.e., when the individual has a plan to either do or not do the act, see Warshaw & Davis, 1985)), although measuring intention in the absence of intention (i.e., when the individual has not decided either way) increases the likelihood of "intention instability" resulting from subjects changing from not having an intention to having one after the measurement of intention but before the performance of the target behavior. Such intention instability is one of the factors that reduces the ability of a measured intention to predict a future behavior (Ajzen & Fishbein, 1975, p. 370). In cases where subjects have not formed an intention for or against a behavior, their attitude is expected to predict the behavior better than their intention (Warshaw & Davis, 1985b).

Intention is theorized to causally mediate between attitude and behavior. Unlike intentions, we theorize that attitudes have been formed regarding the target behavior at the time of measurement, based on Fishbein and Ajzen's (1975) observation that beliefs are generally formed rapidly in response to stimuli (e.g., p. 411-509) and that "as a person forms beliefs about an object, he automatically and simultaneously acquires an attitude toward that object" (p. 216). In the present context, therefore, we are evaluating the relationship between an individual's attitude and future behavior, without explicitly modeling the mediating role of intention. Parenthetically, attitude researchers generally studied the direct attitude-behavior relationship prior to the introduction of the Fishbein intention model (e.g., Wicker, 1969), and continue to pursue a better understanding of moderating variables and conditions under which attitudes are predictive of behavior (e.g., Davidson & Jaccard, 1979; Fazio & Zanna, 1978). One of the most significant contributions in this area was Ajzen and Fishbein's (1977) review of literature on the attitude-behavior relationship wherein they found that "strong attitude-behavior relations are obtained only under high correspondence between at least the target and action elements of the attitudinal and behavioral entities (p. 888)." As discussed previously, the attitude construct in the present model corresponds in specificity with respect to target, action, context, and time frame elements.

Related models. There are a small number of models in the marketing literature which are similar in form and structure to, and which provide further theoretical foundation for, the present model. Hauser and Urban (1977) presented a model which linked objective characteristics of product choice alternatives to choice behavior via perceptual and preferential constructs. They presented data from a study of the design of a new health maintenance organization (HMO). First, the salient set of health care alternatives were "evoked" from subjects. Next, a set of 16 important product attribute items were elicited from subjects using Kelly's (1955) repertory grid methodology on the evoked set. Ratings of the existing alternatives on the attributes were factor analyzed yielding four underlying dimensions. The importance of the perceptual dimensions on preference was modeled using monotone regression and von Neumann-Morganstern utility theory. The linkage between preference and choice intent was modeled using a multinomial logit model of stochastic choice. In addition, they analyzed segmentation and aggregate market share for the product alternatives. Extensions and applications of this basic approach have also been presented (e.g., Hauser & Simmie, 1981; Tybout & Hauser, 1981).

Holbrook (1981) also presents a feature-perception-preference model. The objective features of a piece of piano music were factorially varied. Thirty-eight semantic differential attitude pairs were factor analyzed to reveal five key perceptual dimensions. Affect was measured using evaluative semantic differential items. Regression-based recursive path analysis was used to assess the causal relationships of features on perceptions and perceptions on affect.

Four of the perceptual dimensions were found to be significant determinants of affect, and with one minor exception, mediated between features and affect.

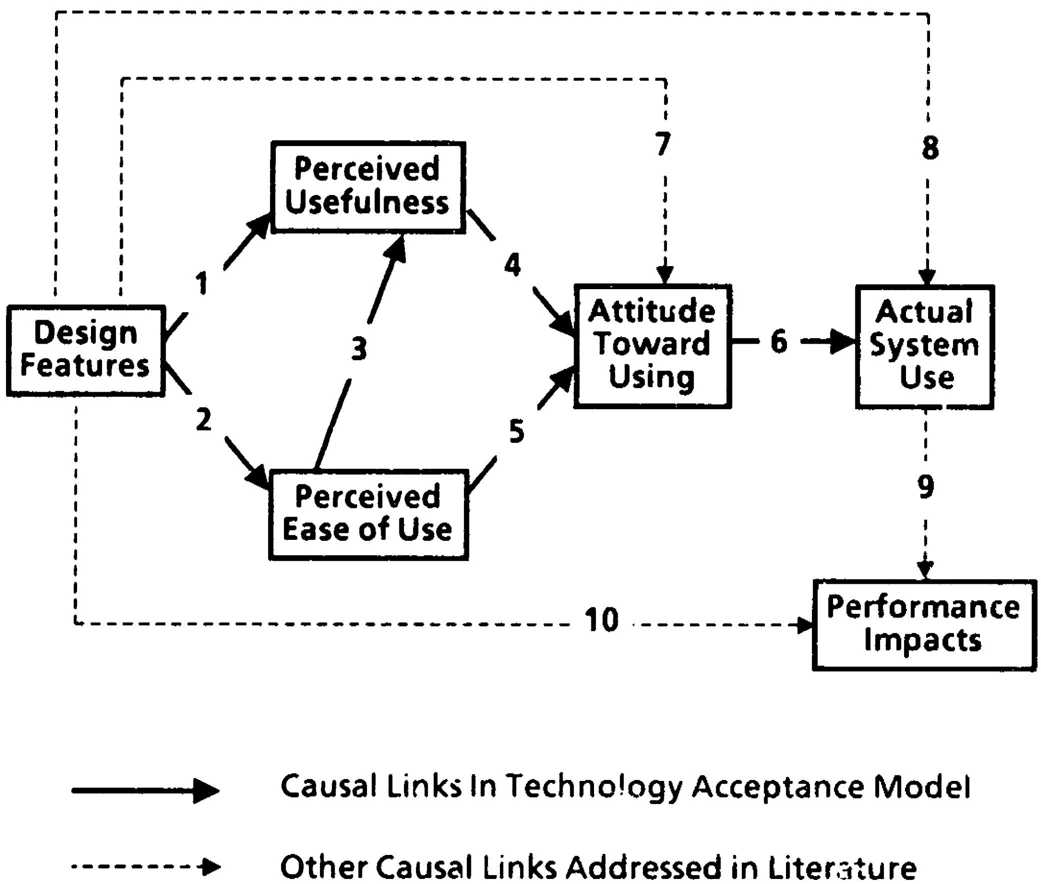
Although the above-mentioned models are similar to the present technology acceptance model, there are a number of minor differences. Whereas the preference and affective constructs in these models appear to be object-based (A_o), the present research employed behavior-based (A_{act}) affective constructs, which Ajzen and Fishbein (1977) argue are more strongly linked to, and hence more predictive of, the target behavior. Correspondingly, the above-mentioned models employ perceived attributes of the object, as opposed to perceived consequences of performing a behavior relative to the construct. Although Hauser and Urban model choice intent, intention is not included in the present model for reasons discussed above. Whereas the above approaches identify perceptual dimensions via factor analysis, the present model posits a priori belief constructs based on theoretical considerations and prior research. Finally, whereas the above models are proffered as general product design methodologies, the present model is specifically tailored for modeling the user acceptance of computer-based information systems in organizational settings.

CHAPTER 3. LITERATURE REVIEW

The purpose of this chapter is to analyze the relationship between the proposed technology acceptance model and the existing empirical literature on the design and implementation of computer-based informations systems. By building the proposed model upon existing work in the field, the present research attempts to follow the advice of Keen (1980) who argues for the importance of establishing a "cumulative tradition" for MIS research. The literature reviewed is drawn primarily from two fields: Management Information Systems (MIS) and Human Factors. The MIS literature is further broken down into lab vs. field studies due to the important differences between these two sub-areas from the standpoint of which causal relationships they have addressed. The objective of the analysis is to: (1) gain an understanding of the existing state of theory and research pertaining to user acceptance processes; (2) identify existing evidence that may lend support to the proposed model structure, and (3) determine the extent to which the proposed model goes beyond existing research.

The method of reviewing the literature used is to analyze which causal relationships prior studies have addressed in comparison to the proposed model. This analysis is carried out with reference to Figure 3.1, which depicts the TAM relationships along with additional links (numbererd 1-10) which have been studied in the literature. Although this set of articles is not exhaustive, it provides a representative picture of research in the area. Articles for the review were identified by searching through key MIS journals, by using reference lists of already obtained articles, and by consulting periodical indices. Articles not dealing with at least one of the causal relationships in Figure 3.1 were not included in the analysis. In order to classify articles in a coherent manner, fairly

Figure 3.1. Technology Acceptance Model Extended for Literature Review



broad interpretations of the definitions of the TAM constructs were used. This made it possible to accommodate the great diversity of existing approaches using a simple taxonomy. In particular, studies coded as having addressed the attitude construct encompassed a diverse set of operational and conceptual definitions of attitude, including perceived attributes, perceived consequences, social influences, attitudes toward the object (e.g., system), satisfaction, etcetera. This reflects the broad way in which the attitude label has been used in the field (Swanson, 1982). Constructs referring to such things as "relevance to job" or "importance" were coded as being equivalent to the perceived usefulness variable due to their conceptual similarity. Similar flexibility in interpretation was used in interpreting whether or not a variable was tapping ease of use. If the definitions of attitude toward using, perceived usefulness and perceived ease of use as defined in the present research were strictly adhered to for this review, most of the studies would not comply, thus defeating our purpose. Design features, use and performance were much more straightforward to classify. Thus, some judgement on the part of the researcher was necessary in classifying studies, with the objective of being flexible while at the same time remaining close to the spirit of the conceptual variables being modeled.

MIS Lab Experiments

MIS lab experiments (see Table 3.1) have typically employed multi-time period decision-making simulations using student subjects. The Minnesota Experiments (for review, see Dickson, Senn & Chervany, 1977) typify this paradigm. The major design features addressed have been information format (tabular vs. graphical information displays, raw vs. statistically summarized data), type of decision support tool, and the like. Dependent variables are typically profit and expense performance within the decision simulation, although information usage and perceptual and attitudinal variables received scattered attention. In addition, several of these studies measure cognitive style and include it as one of the independent variables. Referring to Table 3.1, we find that MIS lab experiments have focused heavily on the design features - performance relationship (link 10). As pointed out previously, however, performance impacts will not be derived if the user does not use the system (in decision simulations, the user is generally required to use the system due to the nature of the experiment).

Table 3.1 Relationship between Prior MIS Lab Experiments and Proposed Model

MIS Lab Experiments	Causal Relationships									
	Within TAM						Outside TAM			
	1	2	3	4	5	6	7	8	9	10
Benbasat & Schroeder 1977								●		●
Chervany & Dickson 1974										●
DeSanctis 1983						●				
Lucas 1981	●						●			●
Lucas & Neilson 1980										●
Lusk & Kersnick 1979										●
Remus 1984										●
Zmud 1978	●	●								
Zmud, Blocher & Moffie 1983										●
Technology Acceptance Model	●	●	●	●	●	●				

Benbasat and Schroeder (1977) used a decision simulation experiment to study the effects of report format (tabular vs. graphical), availability of decision aids, exception reporting (vs. none), number of reports available, decision making style and knowledge of functional area on cost performance, time performance (both link 10) and number of reports requested (a usage metric, link 8). At each of 10 decision points in the inventory/production simulation, the 32 student subjects made decisions regarding order points, order quantities and production schedules, and were allowed to purchase information reports. Of the 18 causal relationships tested (6 independent times 3 dependent variables), 3 were found significant at the .05 level: the effect of decision aids on cost performance, the effect of decision aids on time performance, and the effect of number of reports available on number of reports requested.

Chervany and Dickson (1974) evaluated the impact of raw versus statistically summarized information reports on production costs, decision time and decision confidence for 22 graduate business administration students in a production/inventory simulation. Subjects were asked to assume the role of a production manager responsible for making production schedules, work force changes and material orders at weekly decision points so as to minimize plant costs. Subjects using the summarized reports exhibited better cost performance, lower decision confidence (both non-significant) and took significantly longer to make their decisions relative to those using the raw data reports. This study was therefore concerned with the features-performance link (10).

DeSanctis (1983) assessed the ability of an expectancy theory model to predict use of a decision support system in a laboratory setting. The model was formulated in accordance with the Vroom (1964) tradition of work motivation models. This study was non-experimental in the sense that no independent variables were manipulated. Eighty-eight undergraduate subjects performing a business simulation were given the opportunity to voluntarily use a decision support system. DeSanctis tested the following model:

$$\text{Use} = f(\text{motivation})$$

$$M_i = \Sigma_{j=1,n} \; [(E_{ij} * (\Sigma_{k=1,m} \; V_k I_{jk})]$$

where

M_i = the motivational force to use a DSS at some level i

E_{ij} = the expectancy that a particular level of use (level i) will result in a certain quality of decision making (j)

V_k = the valence, or attractiveness, of outcome k

I_{jk} = the perceived probability that a decision of quality j will result in attainment of outcome k

Use-performance and performance-outcome expectancies, and outcome valences were measured and combined to form a measure of "motivational force." Forces from one time period of the simulation were correlated with actual system use for the subsequent time period. The force-use correlations were small but significant, ranging in value from .042-.239. Although "motivational force" as measured by DeSanctis is largely cognitive in nature, and hence not strictly equivalent to attitude toward using, the theoretical similarity of these variables as behavioral determinants suggest that DeSanctis' study be coded as addressing the link between attitude toward using and use (link 6) on Table 3.1.

Lucas (1981) assessed the influence of type of terminal output (graphs vs. tables and CRT vs. Hardcopy) on decision performance (link 10), self-reported usefulness (link 1), and satisfaction using a simulated whisky importing firm. Subjects were 119 summer executive program students. The study found that subjects using tabular hardcopy terminals performed significantly better, rated their output as more useful and exhibited better problem understanding than those using tabular output on a CRT. In comparing graphics to tabular output on a CRT, graphics groups rated the output as less useful, exhibited greater problem understanding and showed no difference in decision performance. Subjects receiving both graphical and tabular outputs reported significantly greater usefulness than those receiving graphics alone. This experiment also found that analytic versus heuristic decision style interacts with tabular versus graphic information format in influencing problem understanding.

Lucas and Neilson (1980) studied the impact of CRT versus teletype display devices, basic versus graphic report format, and amount of information upon performance and learning (link 10) in a logistics management simulation wherein 36 MBA students, 36 practicing industrial engineers and 42 senior executive program students each operated a firm competing in a simulated industry. Warehousing and shipping decisions were made over 20 simulated weeks. The study found CRT to be superior to teletype for decision performance, although no significant differences for graphics was found. The study also found that additional information will not necessarily result in better performance or learning.

Lusk and Kersnick (1979) conducted a study to determine the impact of report complexity and psychological type on task performance. The experimental task used in this study is quite unlike the production/inventory decision tasks of other studies reviewed in this section. Undergraduate subjects (n = 219) were asked to answer twenty questions using information contained in one of five reports (the experimental treatments) which differed in terms of how the data were expressed: in raw form; as frequencies; as percentages, or in graphical form. The accuracy of their answers were used as a performance criterion (link 10). The five report formats were rated for their complexity in a pre-test. Although no statistically significant performance differences were found between pairs of reports that were adjacent to one another in rank-order complexity, the two less complex reports exhibited significantly greater performance when compared to the three more complex reports. Further, for three of the five report types, analytic psychological types out-performed heuristic types.

Remus (1984) addressed the impact of graphical vs. tabular data formats on the decision performance of 53 undergraduate business majors in a simulated production scheduling context. The use of the tabular display led to lower (although not significantly) decision costs when compared to the graphical display (link 10).

Zmud (1978) studied the impact of three report formats (graphical, tabular, and bar chart) on several perceptual items rated by 35 student subjects. The goal of this study was to derive the dimensionality of the concept of information. The subject's tasks were to evaluate the information content of the 3 reports. This study differs in form and content from the typical decision simulation study. Though neither the number of items nor their identity were specified in the article, the subject ratings were factor analyzed, revealing several dimensions of information. Of the eight dimensions that were derived from the factor analysis, one related to ease of use ("readable") and another related to usefulness ("relevant"). Therefore, this study addressed links 1 and 2.

Zmud, Blocher and Moffie (1983) undertook an experiment to assess the impact of task complexity and color graphic versus black and white tabular report format on decision accuracy, decision confidence and subject learning. Fifty-one professional internal auditors performed risk assessment of simulated invoices. Task complexity was found to be a major determinant of decision confidence and accuracy. Although report format did not exhibit a main effect on decision accuracy (link 10), it did exhibit a significant interaction with task complexity.

MIS Field Studies

MIS field studies have focused largely on the determinants of successful

Table 3.2 Relationship Between Prior MIS Field Studies and Proposed Model

MIS Field Studies	Causal Relationships									
	Within TAM						Outside TAM			
	1	2	3	4	5	6	7	8	9	10
Barber & Lucas 1983							●			●
Fudge & Lodish 1977									●	
Fuerst & Cheney 1982						●				
Ginzberg 1981				●		●				
Ives, Olson & Baroudi 1983				●	●					
King & Epstein 1983				●	●					
Lucas 1975						●			●	
Lucas 1978						●				
Maish 1979						●				
Robey 1979				●		●				
Robey & Zeller 1978				●		●				
Schewe 1976					●	●				
Schultz & Slevin 1975				●						
Swanson 1974						●				
Technology Acceptance Model	●	●	●	●	●	●				

organizational implementation of systems, and have tended to give greater attention to motivational phenomena such as attitudes and perceptions. However, since these are typically organizationally-based field studies the researchers are usually not at liberty to experimentally manipulate the characteristics of the system under investigation. As a consequence, the linkage between user motivation and design features (links 1 and 2) has received much less emphasis by this camp (see Table 3.2).

Barber and Lucas (1983) examined the impact of system response time on CRT operator productivity and job satisfaction. Included in the set of scales used to assess job satisfaction was a measure of system satisfaction ("I like the terminal"). This variable is similar to the TAM attitude toward using construct. Thus, these authors addressed link 7, although the impact of response time on system satisfaction was not significant. Response time had a strong impact on productivity, however (link 10).

Fudge and Lodish (1977) used a quasi-experimental design to show that the use of an interactive management science model had a significant effect on the sales performance of salespeople relative to those not using the model (link 9). Ten matched pairs of salesmen participated in the study, with one from each pair receiving the model to use, the other serving as a control.

Fuerst and Cheney (1982) conducted a study of factors affecting self-reported utilization of decision support systems (DSS) in the oil industry. Sixty-four DSS users from eight large international oil companies completed the questionnaire. The authors found that training, perceived accuracy, experience and perceived relevancy were all significantly related to self-reported usage of the DSS. For the purposes of this review, the perceptual variables were regarded as attitudinal measures, which suggests that this study was addressing link 6.

Ginzberg (1981) assessed the ability to predict acceptance and utilization of a portfolio management system based on the realism of user's pre-implementation expectations in a longitudinal field study involving 44 portfolio managers. Expectation areas studied included: reasons for developing the system, importance of the system, expected usage mode, expected impacts of the system and system evaluation criteria. Realism of expectations was defined as the difference between the expectations of the users and those of user management and system developers. These were measured prior to implementation. Also measured prior to implementation were a set of pre-implementation attitude measures, including: importance of the system, value of the system, probability that the system will be a success, attitudes toward change and to scientific management approaches, perceived management support and user involvement. Subsequent usage was measured objectively for each user via three measures obtained from system logs. Post-implementation user attitudes were measured using 5 subjective outcome ratings including: perceived value, level of use, likelihood that the system is a success and satisfaction with the system. Ginzberg found that pre-implementation perceptions of importance and value both correlated significantly with post-implementation satisfaction ($r = .31$ and .45 respectively). Since perceived importance and value are similar to usefulness, these relationships were coded as link 4. Realism of expectations was treated as a general attitudinal construct, and was found to correlate weakly with usage ($r = .224$, link 6).

Ives, Olson and Baroudi (1983) performed further psychometric analyses of Bailey and Pearson's (1983) user satisfaction scale. A sample of 200 managers completed 2 separately-mailed questionnaires. The second questionnaire obtained a separate overall measure of satisfaction for assessing predictive validity (correlation between the 2 satisfaction measures was .55). Inter-item reliabilities were found to be above .9 for 30 out of 38 scales. A factor analysis revealed 5 factors underlying the satisfaction index. The authors proposed an abbreviated version of the satisfaction scale. They computed the correlation between each of the 38 scales on an overall scale. Three of the scales were similar to perceived usefulness (perceived utility, $r = .67$; relevancy of output, $r = .77$; and job effects of computer support, $r = .74$) and three were similar to perceived ease of use (understanding of systems, $r = .63$; feeling of control, $r = .68$; and flexibility of systems, $r = .77$). Viewing the overall index as one's attitude toward using, this study addressed links 4 and 5.

King and Epstein (1983) tested the following multiattribute model of information systems value:

$$V = \Sigma_{i = 1,10} W_i * V(a_i)$$

V = overall systems value

W_i = self-reported importance weight for attribute i

$V(a_i)$ = self-reported evaluations for each level of each attribute

Ten attributes were selected to represent information systems: reporting cycle, sufficiency, understandability, freedom from bias, reporting delay, reliability, decision relevance, cost efficiency, comparability, and quantitativeness. The authors measured the importance weights and attribute evaluations from each subject, and then measured the subjects overall evaluation for each of several system profiles expressed in terms of the 10 attributes. For three different groups of managers, the correlation between overall value as predicted by the model versus directly rated was .721, .918 and .856 respectively. The overall value variable is similar to attitude toward using, and of the 10 attributes, one relates to usefulness ("decision relevance") and one relates to ease of use ("understandability"). Therefore, this study was coded as analyzing relationships 4 and 5. Unfortunately, neither the importance weights nor the correlation between each individual attribute and overall value was reported.

Lucas (1975) proposed and tested a model of the determinants of information systems use and performance. Use was hypothesized to be influenced by: performance, situational factors, personal factors, decision style, and attitudes and perceptions toward the system. Performance was hypothesized to be a function of situational and personal factors, decision style and use. Empirical data was gathered from 316 salesman and 82 account executives regarding their use of a sales information system. Performance data came from computer files, as did 2 of the 4 situational factors, the remainder of the data coming from a questionnaire. Stepwise regression analyses provided support for the attitude-use relationship (link 6), and mixed support for the use-performance relationships (link 9). Perceived output quality was one of the attitudinal variables found to be linked with several different usage variables (standardized regression weights varying from .22 to .38).

Lucas (1978) examined the ability of a series of attitude measures to explain the degree of use of a medical research information system. Usage was assessed with 15 different measures (7 based on a questionnaire, 8 based on system monitors). The attitudinal constructs were almost all significantly correlated with use.

Maish (1979) conducted an attitude survey involving 62 respondents from four Federal agencies. The questionaire contained attitude questions covering a range of topics including management support, quality of staff, access, and information formats. A 4-item self-reported usage index was also included.

Several of the attitude scales were significantly related to self-reported use (link 6), with pearson correlations ranging from .10 to .54.

Robey (1979) used Schultz and Slevin's (1975) attitude instrument to explain actual use of a computerized record-keeping system by 66 salespeople. Two indicators of actual use were obtained were obtained from company records (these correlated .97). Of the 7 Likert subscales making up Schultz and Slevin's instrument, 2 were eliminated due to low internal reliabilities. The remaining 5 subscales were all found to be significantly correlated with both of the usage measures (link 6). Of those, the "performance" subscale was most highly correlated with use (Spearman correlations = .79 & .76 respectively for the two use measures). Although performance is similar to perceived usefulness, no separate attitude measure was taken, and hence performance was regarded as an attitudinal variable.

Robey and Zeller (1978) used the Schultz and Slevin instrument to diagnose the reasons why one department in a company adopted and used an information system, while a similar department in the same company rejected the same system. The Schultz and Slevin instrument was administered to both departments. Across the 7 subscales of the instrument, 2 were significantly different between departments: performance, and urgency (link 6).

Schewe (1976) conducted a survey of 79 computer users from 10 food processing firms. Usage was measured via a self-report of the "number of monthly requests that a manager/system user made for additional information". A wide range of attitude and belief measures were taken. The results indicate that attitudes show no relationship with usage (link 6). This may be attributable to the particular usage measure employed: "the number of monthly requests that a manager/system user made for additional information" (p. 584). The subjects may have made substantial use of the system without making such requests. Strong relationships were observed between many of the belief items and attitudes, none of which are pertinent to the present analysis.

Schultz and Slevin (1975) developed an instrument for measuring individual attitudes toward OR/MS innovations. Through a review of the literature, they selected an initial set 81 variables thought to be pertinent to "problems of implementation in the organizational environment" (p.156). These were used to form Likert statements worded in terms of "what would happen as a result of the implementation". Eleven semantic differential concepts, also based on the literature, were created and included "as somewhat of an exploratory instrument." A pilot test using 136 MBA students was performed. Responses to the Likert and semantic differential items were factor analyzed, yielding a representation of the responses in terms of 10 factors. The pilot test led to a revised questionnaire containing 67 Likert items and 11 semantic differential items. Five dependent variables were added at this point (probability of self using, probability others will use, probability of success, perceived worth, perceived level of accuracy).

Next, Schultz and Slevin field tested the revised instrument on 94 management personnel regarding a new computer forecasting model that was being implemented in a manufacturing company. A factor analysis of the 67 Likert items yielded 7 factors, labeled: performance; interpersonal; changes; goals; support/resistance; client/researcher; and urgency. The performance factor is viewed as the perceived "Effect of model on manager's job performance." This is almost exactly the same as the definition of perceived usefulness in the context of TAM. Also, the "perceived worth" dependent variable was measured on a scale with anchoring adjectives: "not useful at all"; "moderately useful"; and "excellent", thus probably tapping into TAM's "perceived usefulness" construct. The correlation between perceived worth and the performance factor (r = .59) is consistent with this possibility. Of the eleven semantic differential concepts, concepts 1 and 7 (chance of success & importance) may be viewed as evaluative of the system per se. Moreover, the 10 adjective pairs used to measure each semantic differential concept were those known to load significantly on the evaluative dimension (Osgood, Suci & Tannenbaum, 1957). Since semantic differential scales employing evaluative adjective pairs represent the most highly reccomended approach for measuring attitudes within the Fishbein paradigm (Ajzen & Fishbein, 80; Fishbein, 64, 67; Fishbein & Ajzen, 75) the ratings of these two concepts can be viewed as an assessment of attitude toward the system (though not necessarily attitude toward using the system). The high observed correlation between these two concept ratings and both the performance factor (r = .56 & .63, respectively), and perceived worth (r = .68 & .68, respectively) can thus be regarded as evidence of the usefulness-attitude relationship (link 4). Additionally, these two semantic differential scores also correlated highly (.53 & .68) with the first of the dependent variables (self-predicted use). This is consistent with the theoretical relationship between attitude and behavioral expectation (Warshaw & Davis, 1985).

Swanson (1974) measured a construct that he called "MIS appreciation" using 16 perceptual items. Due to the evaluative nature of these items, the appreciation construct is similar to attitude. Using a non-parametric Chi-square test, Swanson found appreciation to be strongly related to a system usage indicator called "inquiry involvement" ($p<.01$, link 6).

Human Factors Literature

The Human Factors literature (Table 3.3) has long been concerned with a broad range of design features for several types of systems, especially focusing on text editors and database query languages. These researchers have begun to give increasing attention to perceptual and attitudinal dependent variables.

Bewley, et al. (1983) report on four experiments that were performed as part of the design of the Xerox 8010 'Star' workstation. The first two were concerned with the assignment of cursor control functions to the buttons of a mouse pointing device. The first experiment compared six different approaches with respect to selection time. The insights gained from the first experiment were used to formulate a seventh scheme which was demonstrated via the second experiment to provide even faster pointing time than the first six. Thus, experiments one and two were concerned with the features-performance (10) link. The third experiment compared 4 sets of icons with respect to recognizability via timed tests (link 10) and subject-rated "ease picking out of a crowd" (link 2). Apparently, no significant differences were detected. The fourth experiment compared two versions of a graphics editor interface with respect to illustration time (link 10).

Card, English and Burr (1978) compared the mouse, isometric joystick, step keys and text keys as cursor positioning devices for selecting text on a CRT. The dependent variables were speed of selecting text on a CRT, learning time, and error rate (link 10). The distance to the target, target size, and approach angle were all varied in the experimental design. They found the mouse to be the

Table 3.3 Relationship between Prior Human Factors Experiments and Proposed Model

Human Factors Experiments	Causal Relationships									
	1	2	3	4	5	6	7	8	9	10
Bewley, et al. 1983		●					●			●
Brosey & Shneiderman 1978										●
Card, English & Burr 1978										●
Card, Moran & Newell 1980										●
Good 1982		●					●			●
Gould, Conti & Hovanyecz 1983							●	●	●	●
Greenblatt & Waxman 1978										●
Ledgard et al. 1980							●			●
Lochovsky & Tsichristzis 1977										●
Magers 1983	●	●					●			●
Malone 1981						●	●	●		
Miller 1977	●	●					●			●
Poller & Garter 1983		●								●
Price & Cordova 1983										●
Reisner, et al 1975										●
Roberts & Moran 1983										●
Shneiderman et al. 1977									●	
Thomas & Gould 1975										●
Welty & Stemple 1981										●
Technology Acceptance Model	●	●	●	●	●	●				

superior pointing device and, observing that it's use appears to be governed by Fitt's Law, argued that pointing time using the mouse is close to the theoretical minimum time achievable.

Card, Moran and Newell (1980) formulated and empirically assessed their "keystroke-level model" for predicting user performance times. The model breaks down the time required for an expert user to perform a specified task into its component parts. The task components are determined in part by the characteristics of the editor (e.g. number of keystrokes required to perform a task, type of cursor positioning device) and in part by human capabilities. The time required to perform each of the components, which consist of human mental and motor operations as well as system response time, are estimated using standard data from time and motion studies, empirical laws and other sources. The predicted times based on the model correlated highly ($r = .87$) with actual performance times. In Table 1 this study was coded as having addressed the link between design features (i.e., # keystrokes required) and performance (link10).

Good (1982) compared Etude, an interactive document processing system, with a standard typewriter with regard to training and performance time (link 10), subject anxiety, and 12 semantic differential items. Two of the semantic differential items (unfriendly-friendly; helpful-unhelpful) can be viewed as tapping perceived ease of use (link 2), and at least two of them (awful-nice; unpleasant-pleasant) were related to attitude toward using (link 7), though only the awful-nice score was significantly different across the two test systems.

Gould, Conti and Hovanyecz (1983) performed two experiments comparing several versions of a simulated listening typewriter. In the first experiment, eight different composing methods, created by taking all combinations of two design features: speech mode (isolated vs. consecutive words) and vocabulary size (1000 words vs. unlimited), and composing strategy (draft vs. final version), were compared for the task of writing a letter. Writing the letter by hand was included as an additional method. The dependent variables were composition time, effectiveness of the letter as ranked by judges (both link 10), and preference of using each method relative to writing (link 7). Since using a manual method was compared to the various versions of the system, the experiment was coded as having also addressed the use-performance link (9). Whereas the first experiment aimed to compare the listening typewriter to writing a letter by hand, the second compared it to dictation. Five versions of the listening typewriter, created by varying the design features, were compared to one another and to dictating to a dictation machine and a secretary. Dependent variables were: composition time, proof time, and letter effectiveness (link 10); and preference (links 7). In addition, the second experiment studied the features-use link (8) by asking subjects to select their choice of any version of the listening typewriter, a dictation machine or a secretary for performing a final dictation task.

Ledgard, et al. (1980) compared a natural language syntax to a more traditional notational syntax for an interactive text editor. Dependent variables were objective performance measures such as editing efficiency (link 10), as well as subjective preference ratings (link 7). The natural language interface was strongly favored.

Magers (1983) addressed the effect of on-line help on several performance variables (including task time and errors, link 10) and several perceptual and attitudinal items (link 7), including at least 8 that related to ease of use (link 2) and 1 that related to (command) usefulness (link 1). On-line help had a positive effect on nearly all response variables

Malone (1981) reported three studies related to the design of computer games. The first study was a survey of computer game preferences among elementary schoolchildren. Various features of the games were found to be significantly correlated with preference (link 7). In the second study, six versions of a computer game consisting of various combinations of features were compared in terms of how well they were liked (link 7). Again, significant effects were observed. The third study also addressed the link between design features and how well liked the version was, but included a behavioral usage measure of the amount of time the subjects spent playing the treatment game versus a control game (link 8). The correlation between preference and usage was .30 (link 6).

Miller (1977) studied the impact of baud rate, output rate variability, and output volume of an interactive message retrieval system on user performance (link 10) and an 18-item measure of user satisfaction (link 7). One of the items measured perceived ease of use of the system's commands (link 2) and five of the items formed an index of perceived usefulness (link 1). Interestingly, he found that while baud rate had no effect on attitude or performance, rate variability had highly significant negative effects on performance, ease of use, and usefulness.

Poller and Garter (1983) compared moded versus modeless text editing in terms of perceived ease of use (link 2), speed, and error rate (link 10), finding that while the modeless editor fared better on ease of use criteria, it took longer and was associated with more errors. This is an interesting finding which suggests that perceived ease of use may not neccessarily be related to "actual" ease of use as operationalized via performance measures.

Price and Cordova (1983) studied different configrations of mouse buttons, finding that: "People tended to be faster and more accurate using different buttons than different numbers of clicks" (p. 262) in various pointing tasks (link 10).

Roberts and Moran (1983) report on a comparison between nine existing text editors. The dependent variables were: time to perform basic editing tasks, error time, learning time and an analytical measure of an editor's functionality (defined as the percent of a pre-defined set of tasks that are feasible to do on the editor). The time required to perform editing tasks as predicted by Card, et al.'s (1980) keystroke-level model correlated .9 with actual performance time by expert users. Although the editor's features themselves were not experimentally manipulated, 4 characteristics of the editors were found to be significantly correlated with learning time: number of core commands in editor, number of physical operations per task, number of method chunks per task, and expert time score. The study was therefore coded as having addressed the link between design features and performance (link 10) on Table 1.

Shneiderman, et al. (1977) report on five experiments that compared the use of flowcharting versus not using it in various programming, debugging, and program modification tasks. The dependent variables were all objective performance criteria and, since the independent variable was using vs. not using, as opposed to a comparison of different flowcharting techniques, these studies were coded on Table 3.2 as having addressed the use-performance link (link 9). Surprizingly, no significant effect for flowcharting was found across all five studies.

A series of experiments concerned with the design of database query languages have been published (for a review, see Reisner, 1981). These studies share the characteristic that they focus primarily on the features-performance link (10). Several of the studies also measure the confidence in the correctness of the query, although this construct is not represented in the proposed model of the present research and hence does not have a causal link represented on Figure 3.1. Reisner, Boyce and Chamberlin (1975) compared the SQUARE and SEQUEL query languages. The dependent variables were various measures of correctness and ease of remembering, measured by paper and pencil tests. Thomas and Gould (1975) report on the objectively defined correctness as well as the subjective confidence in the correctness of queries written using the Query by Example language, although no experimental comparison with other languages was performed. Greenblatt and Waxman (1978) compared three relational database query languages in terms of query correctness, query confidence and speed of query writing. Lochovsky and Tsichristzis (1977) compare the hierarchical, relational and network data models while keeping the command language syntax constant. Their dependent variables included coding accuracy, coding time, debug time, query comprehension and query correctness. Brosey and Shneiderman (1978) report on two experiments analyzing the effect of data model (relational vs. hierarchical) on comprehension, problem solving and memorization. Welty and Stemple (1981) compared two languages that differed primarily in their degree of procedurality: SQL and TABLET, using query correctness as the evaluative criterion.

Related Studies

In addition to the research studies discussed above, there are a few studies that are pertinent to the present research that do not address one of the specific relationships of Figure 3.1. These are discussed below.

Bailey and Pearson (1983) developed a measure of "computer user satisfaction". They identified a list of 39 factors by generating an initial list of factors from a literature review, and then adding additional factors suggested by data processing professionals who they interviewed. The expanded list was checked for completeness via a critical incidents analysis. Four semantic differential scales with appropriate end-point adjectives were used to measure the subject response to a factor; an importance weight was measured for each factor as well. The responses and importance weights were combined according to the following formula:

$$S_i = \Sigma_{j=1,39} W_{ij} / 4 \ \Sigma_{k=1,4} I_{ijk}$$

S_i = satisfaction score for user i

W_{ij} = Importance weight for factor j

I_{ijk} = subject i's response to item k of factor j

Of the 39 factors addressed by Bailey and Pearson, three appeared to be related to perceived ease of use: "understanding of systems","feeling of control" and "flexibility of systems" ; and three appeared to be related to perceived usefulness: "relevancy", "perceived utility", and "job effects". The questionnaire was completed by 29 subjects, and the resulting data was used to assess the reliability and validity of the instrument. The average reliability for the 4 items for each factor was .93. The relationship between

Gallagher (1974) compared two alternative approaches for assessing the value of information reports: estimated annual dollar value and semantic differential opinions. In a field study of 52 managers, the correlation between these two approaches was .29 ($p < .05$). Many of the 15 semantic differential items bear a resemblance in meaning to the TAM perceived usefulness construct: useful-useless; relevant-irrelevant; important-unimportant; applicable-inapplicable; necessary-unnecessary.

Larcker and Lessig (1980) present research on the preliminary development of an instrument for measuring "perceived usefulness" of information. Based on their review of the literature, the authors propose that perceived usefulness consists of 2 distinct dimensions. The first, called "perceived importance", refers to "the quality that causes a particular information set to acquire relevance to the decision maker, and is a function of whether the information items "are a necessary input for task accomplishment". The second dimension, called "perceived usableness", is a function of whether "the information format is unambiguous, clear, or readable". An open-ended elicitation process was used to generate an initial set of thirteen attributes for the dimensions. Next, the items were rated by 6 judges as either perceived impoortance or perceived usableness. Seven of the items were eliminated, yielding 3 items for each dimension. Then a survey of 29 faculty and graduate students was performed to assess reliability and content validity. Each subject rated 4 different information sets in the context of given decision scenarios. Factor analysis and multitrait-multimethod analysis confirmed that importance and usableness are distinct dimensions. Reliabilities of the three-point scales ranged between .636 and .773.

The two dimensions studied by Larcker and Lessig are similar to the two perceptual dimensions of the proposed technology acceptance model. Perceived importance refers to whether the information is necessary for the task at hand, whereas perceived usefulness refers to the expected impacts on productivity resulting from system use. Thus, while importance refers to the feasibility of doing the task, usefulness is concerned with whether the task that the system performs is an important part of the person's job. Perceived usableness is quite similar to perceived ease of use, both having to do with the amount of effort to use the system (or information set) to perform the target task. The fact that Larcker and Lessig found usableness and importance to be distinct constructs lends support for the representation of perceived usefulness and perceived ease of use as distinct constructs in the technology acceptance model.

Discussion

Looking across the three categories of literature reviewed, we find existing empirical support for all six of TAM's relationships except for the ease of use-usefulness link, which none of the reviewed studies addressed. Significant relationships were found between system characteristics and both perceived usefulness (Lucas, 1981; Magers, 1983; Miller, 1977) and perceived ease of use (Bewley, 1983; Magers, 1983; Miller, 1977; Poller & Garter, 1983). Attitude was significantly effected by both perceived usefulness (Ginzberg, 1981; Ives, Olson & Baroudi, 1983; Schultz & Slevin, 1975), and perceived ease of use (Ives, Olson & Baroudi, 1983; Schewe, 1976). Although a wide array of attitudinal and usage measures were employed, many researchers observed a significant attitude-usage relationship (Fuerst & Cheney, 1982; Lucas, 1975; Lucas, 1978; Maish, 1979; Robey, 1979; Robey & Zeller, 1978; Swanson, 1974). Researchers who found a nonsignificant relationship between attitude and usage appeared to be using either new unproven attitude measurement procedures (e.g., "force", DeSanctis, 1983; "realism of expectations" Ginzberg, 1976) or questionable usage operationalizations (e.g., Schewe's (1976) measure of requests for additional information).

Thus, non-trivial empirical support already exists for the five of the six individual causal relationships reflected in TAM. At the same time, none of the reviewed studies have dealt with all six of the TAM relationships. In this sense, TAM tends to integrate previous findings, yielding a more complete specification than previously approaches. For example, looking at the MIS laboratory studies (Table 3.1) and Human Factors studies (Table 3.3), we observe that these laboratory-oriented studies have been concerned with the effects of design variables. The majority of these studies have traditionally used some form of performance criterion as the dependent variable, but have increasingly paid attention to attitudinal and perceptual variables. However, they have tended not to address the relationships between perceptions, attitudes, and usage behavior. In contrast, MIS field studies have given most of their attention to modeling the perceptual and attitudinal determinants of usage behavior, while generally ignoring one of the key managerially controllable variables affecting these behavioral determinants: system characteristics. The present research adopts the position that these approaches are complementary, and that an integrative model encompassing the effect of design features on perceptions on the one hand, and the effects of perceptions on attitudes and behavior on the other, is a logical next step toward advancing the frontier of theory pertaining to user acceptance processes.

CHAPTER 4. SURVEY

Objectives

There are two major objectives for the survey reported in this chapter: (1) to empirically test the hypothesized causal structure of the proposed technology acceptance model (TAM), and (2) to develop and validate scales for measuring the TAM theoretical constructs. These are each discussed in greater detail below.

Model Testing Process and Rationale

The survey data will be used to test hypotheses regarding the causal structure of the proposed Technology Acceptance Model. The equations defining the proposed model are given below:

$$\text{EOU} = \beta_1 \text{ System} + e$$

$$\text{USEF} = \beta_1 \text{ System} + \beta_2 \text{ EOU} + e$$

$$\text{ATT} = \beta_1 \text{ EOU} + \beta_2 \text{ USEF} + e$$

$$\text{USE} = \beta_1 \text{ ATT} + e$$

The model will be tested according to the structural equation modeling paradigm defined by Duncan (1975), Land (1973) and Simon (1954). Within this paradigm, the proposed technology acceptance model is "recursive" in that "no two variables are reciprocally related in such a way that each affects and depends on the other, and no variable 'feeds back' upon itself through any indirect concatenation of causal linkages" (Duncan, 1975, p. 25). Land (1973) shows that recursive models are identifiable and that ordinary least squares (OLS) regression applied to each equation provides optimal (minimum variance linear unbiased) parameter estimates. The structural coefficients to be estimated are given in Table 4.1. Consistent with the structure of recursive

Table 4.1. TAM Constrained and Unconstrained Parameters

Dependent Variable	Independent Variable				
	SYS	EOU	USEF	ATT	USE
System (SYS)	---	0	0	0	0
Perceived Ease of Use (EOU)	β	---	0	0	0
Perceived Usefulness (USEF)	β	β	---	0	0
Attitude toward Using (ATT)	0	β	β	---	0
Actual System Use (USE)	0	0	β	β	---

models, the coefficients above the diagonal are restricted to be zero. The parameters specified as β's are hypothesized to be significant according to the proposed model. The below-diagonal coefficients designated as zero in Table 4.1 correspond to relationships which are theorized to be non-significant. The non-significance of these relationships are hypotheses to be tested. Hence, the model testing analysis can be logically broken into two components. First, the statistical significance of those causal relationships hypothesized to be significant will be tested. These tests are expressed in hypotheses 1-6 below and will be assessed using the t-statistic corresponding to each estimated parameter. Second, the data will be used to assess whether the causal relationships hypothesized not to exist are statistically insignificant. These tests are expressed in hypotheses 6 and 7 below. Hierarchical regression and associated F-tests of the significance of the increase in R^2 due to the additional variables will be used for these hypotheses.

H1: Attitude toward using will have a significant effect on actual system use.

H2: Perceived usefulness will have a significant effect on attitude toward using, controlling for perceived ease of use.

H3: Perceived ease of use will have a significant effect on attitude toward using, controlling for perceived usefulness.

H4: Perceived ease of use will have a significant effect on perceived usefulness, controlling for system.

H5: System will have a significant effect on perceived usefulness and perceived ease of use.

H6: Perceived usefulness, perceived ease of use, and system will not have significant direct effects on actual system use, controlling for attitude toward using.

H7: System will not have a significant direct effect on attitude toward using, controlling for perceived usefulness and perceived ease of use.

Following Duncan (1975, p. 49) if a coefficient is theorized to be significant, but found to be nonsignificant in the empirical data, its corresponding independent variable will be left in the regression. To assume that the relationship is non-existent may be an instance of falsely accepting the null hypothesis, which could arise if a true influence exists but is too small to be detected by the statistical tests. To inappropriately remove the non-significant variable may lead to inconsistent estimates of the coefficients of the remaining variables. Similarly, if a relationship hypothesized to be insignificant is found significant, the corresponding independent variable should be included in the regression. Such a finding would be suggestive of model misspecification, and to omit the variable in question may distort estimates of other relationships (Pindyck & Rubenfeld, 1981). Theoretical implications of such unexpected findings should be considered.

In addition to testing for the significance vs. non-significance of the hypothesized relationships, the survey data will also be used to estimate the magnitudes of the causal parameters. The estimates will be the standardized partial regression coefficients, and will be expressed both as point and confidence interval estimates.

Measure Development Process and Rationale

The survey, and the preliminary procedures leading up to it, will be used to develop reliable and valid measures for the theoretical variables of the proposed model. There are two key reasons to develop valid and reliable measures:

(1) To support theoretical research. The technology acceptance model is intended to provide a valid theoretical explanation of what motivates people to use computer systems. Much of the evidence to be used in establishing the validity of TAM is to be obtained by comparing measured observations with theoretically predicted patterns. To the extent that the measures used are flawed, observed relationships would provide a distorted view of the underlying theoretical processes, reducing the likelihood of correct inferences. Thus, the present research, as well as future research that may attempt to revise or extend the proposed model, will benefit from the existence of reliable, valid measures for the constructs being modeled.

(2) To support applied research. The measures are intended to form a central element of an applied user acceptance testing methodology. Unreliable and/or invalid measures could seriously mislead designers regarding the effects of their design choices on user behavior.

Reliability

Reliability refers to the extent to which a measurement item (question) is free of random error (e.g,. Nunnally, 1978, p. 191). The following true-score model is frequently used to conceptualize the role of random error in a measure:

$$X_{ij} = T_{ij} + e_{ij}$$

X_{ij} = observed score from subject i on item j

T_{ij} = true score for subject i on item j

e_{ij} = random error for subject i on item j

Reliability is generally defined as the proportion of variance in the observed score X_{ij} that is due to the true score T_{ij}, or $\sigma^2{}_t / \sigma^2{}_x$. As the amount of random error in a measure increases, its reliability diminishes. Unreliable measures create difficulties for statistical analyses in which they are used. For comparisons between mean values, unreliability inflates the standard errors of the estimated means and thereby increases the likelihood of committing Type II error. Further, unreliability attenuates estimates of correlation and regression coefficients relative to what their true value would be with error-free measures. Although it is generally impossible to completely eliminate random measurement error, it is possible to substantially reduce the error so as to minimize it's effect on statistical tests. We will employ a target reliability level of .80 based on Nunnally's (1978) suggestion that: "For basic research, it can be argued that increasing reliabilities much beyond .80 is often wasteful of time and funds. At that level correlations are attenuated very little by measurement error" (p. 245).

Construct Validity

Although the construct validity of a measurement scale has been defined in a variety of ways in the psychometric literature (e.g., Bohrnstedt, 1970; Cronbach & Meehl, 1955; Nunnally, 1978), a predominant perspective views a measure's construct validity as the degree to which the measure's true score corresponds to the conceptual variable that the measure is intended to operationalize (e.g., Cook and Campbell, 1979, p. 59; Fishbein & Ajzen, 1975, p. 108). Whereas reliability is concerned with the amount of random variance in an observed score, construct validity is concerned with the degree to which the systematic variance in a score corresponds to the target construct.

Invalidity may come about in two conceptually distinct ways. First, a measure's true score component may tap some alternative theoretical variable other than the one intended. The correspondence between a measurement scale and the the theoretical variable of interest is sometimes referred to as "content validity" (e.g., Bohrnstedt, 1970; Nunnally, 1978). If a measure lacking content validity is employed, researchers may incorrectly interpret the resulting data in terms of the theoretical variable that was intended by the measure, rather than the variable that was actually measured. This obviously increases the likelihood of false theoretical inferences. Second, methodological artifacts unrelated to the target theoretical variable, such as individual differences in response set (e.g., Campbell, Siegman, & Rees, 1967; Silk, 1971), may comprise part of the systematic variance in a measure. This type of invalidity may be a source of spurious covariation between variables whose measures are affected similarly by such methodological artifacts. The resulting data may overstate the magnitude of the true underlying relationship. This source of invalidity is sometimes refered to as "shared method variance" (Campbell & Fiske, 1959, p. 85). In the present research, different techniques will be employed to deal specifically with each of these potential sources of measure invalidity.

Multi-item Scales

The present research will employ multi-item measurement scales. Whereas single-item scales tend to be invalid and unreliable, possessing a high degree of irrelevant content along with the target content (Cook & Campbell, 1979, p. 65), the use of multi-item scales tends to allow the irrelevancies of individual items cancel out, increasing reliability and validity. The primary method for increasing the reliability of a scale is to increase the number of items (Nunnally, 1978, p. 243). The individual items will use semantic differential and Likert-type rating formats. These types of items have traditionally been used in attitude scaling, and are the ones recommended by Ajzen and Fishbein (1980) for operationalizing beliefs and attitudes. Substantial experience in applying items such as these has shown that they are generally capable of attaining high levels of reliability and validity (Fishbein & Raven, 1962; Jaccard, Weber & Lundmark, 1975; Ostrom, 1969; Robinson & Shaver, 1969; Shaw & Wright, 1967). Moreover, they are quite easy to use by non-experts, making them suitable for the applied user acceptance testing context in which the model is intended to be used.

Existing Scales

A logical first step in the process of defining scales for measuring the TAM variables is to scan the literature for existing scales that meet the reliability and validity criteria. In the case of attitude toward using, standard, validated, multi-item attitude toward behavior scales are available in the psychology literature (e.g., Ajzen & Fishbein, 1980, Appendix A). Standard measures employ 7-point rating scale formats anchored with evaluative semantic differential (Osgood, Suci & Tannenbaum, 1957) adjective pairs (such as 'good-bad'), and typically exhibit reliability values in the desired range (e.g., Bagozzi, 1981; Fishbein & Raven, 1962). These standard scales are readily adapted to the present context by specifying the desired target (system), behavior (using the system), context (in your job), and time frame (unspecified future). Four to five items are typically employed in order to assure the desired psychometric properties. These standard scales will be used in the present research to measure attitude toward using for TAM.

No existing validated, multi-item scales with the desired reliability of .80 were found for perceived usefulness or perceived ease of use, however. The Fishbein paradigm (e.g., Ajzen & Fishbein, 1980, Appendix A) provides a recommended format for beliefs once they have been specified, although it does not furnish complete scales for specific belief variables. The prior Management Information Systems and Human Factors literature discussed in Chapter 3 was reviewed for existing scales meeting the specified requirements. The majority of studies measuring usefulness or ease of use either employed single-item scales or failed to report the psychometric characteristics of the multi-item scales used. The remaining candidate multi-item scales exhibited reliability below the desired level, were unvalidated, or both. Robey (1979) employed Schultz and Slevin's (1975) instrument, which contains a factor called "performance" that is similar to perceived usefulness. He found a Cronbach Alpha reliability of .81 for the scale, although the original instrument was non-validated, having been developed via exploratory factor analysis, and the performance scale contained items relating to "performance visibility" as well. Larcker and Lessig (1980) did perform a content analysis validation of their 3-item scales of "usableness" and "importance" but the reliabilities fell short of our desired level (.64-.77). Ginzberg's (1981) 2-item "importance" scale achieved a reliability of .59. Bailey and Pearson's (1983) instrument contained three 4-item semantic differential scales of usefulness-like factors ("relevance", "perceived utility" and "job effects") and four 4-item scales of factors that are similar to ease of use ("error recovery", "understanding of systems", "feeling of control", and "flexibility of systems"). However, the definitions given to respondents for each of these factors depart considerably from the conceptual definitions of usefulness and ease of use in the present research. Bailey and Pearson (1983) performed a content analysis validation, although they did it from the standpoint of these factors as measures of "computer user satisfaction" as opposed to usefulness and ease of use per se. Miller (1977) did not present evidence of reliability or validity for his 3-item ease of use and usefulness scales, nor did Schewe (1976) give such evidence related to his 3-item ease of use scale.

Given the lack of sufficiently reliable and valid scales in the existing literature, new scales for perceived usefulness and perceived ease of use will be developed. As will be discussed shortly, the scales found in the existing literature will be used as a source of items for constructing the new scales. Consistent with Ajzen & Fishbein (1980), perceptions will be measured using

Likert-type ('agree-disagree') rating formats. In the survey reported below, the extent of agreement with belief statements is measured using 7-point "circle the number" rating scale formats.

Measure Development Process

The process used in the present research for developing usefulness and ease of use scales was designed to address the three key psychometric properties identified above: reliability, content validity and common method variance. First, an initial pool of candidate items was generated for each construct based on existing MIS and Human Factors literature. Next, pretest interviews were conducted in order to perform a content analysis of the items. The item generation and pretest steps were performed in order to increase the content validity of the measures, and are motivated by the "domain sampling" model of test construction, which is discussed in the following section. The survey provides the data needed to assess reliability and convergent and discriminant validity. Cronbach's (1951) alpha reliability coefficient will be computed from the survey data. Campbell & Fiske's (1959) multitrait-multimethod technique will be applied to the survey data, which will provide circumstantial evidence of content validity and will permit an assessment of the extent of common method variance in the measures.

Measure Development and Pretesting

Measurement Item Generation

The first step in the measure development process is to identify an initial set of measurement items as candidates for the final usefulness and ease of use scales. The candidate items will be derived from published articles that have discussed or attempted to measure the target constructs. As discussed above, the process used for generating items aims to ensure that the items possess content validity, which is defined as "the degree that the score or scale being used represents the concept about which generalizations are to be made" (Bohrnstedt, 1970, p. 91). As Nunnally (1978, 9. 258) points out in his discussion of content validity: "Rather than test the validity of measures after they have been constructed, one should ensure the validity by the plan and procedures for construction." In order to explain why generating items from the existing literature is expected to increase the content validity of the resulting measures, we now introduce the "domain sampling model" of measure construction, which psychometricians frequently employ as a conceptual tool to guide the measure development process (e.g., Bohrnstedt, 1970, p. 92; Nunnally, 1978, p. 193). Under the domain sampling model, there is assumed to be a universe or domain of content corresponding to each variable one is interested in measuring. Under this model, the optimal way to develop a scale would be to specify the domain of content corresponding to the target construct and then randomly sample items from the domain. The mean value of a summative scale composed of such a randomly sampled subset would theoretically provide an unbiased estimate of the mean of all the items in the entire domain (i.e. the universe score), which in turn corresponds to the magnitude of the true underlying construct. However, it is ordinarily not possible to rigorously specify the domain of content corrresponding to psychological constructs (Bohrnstedt, 1970; Cronbach & Meehl, 1955; Nunnally, 1978).

Since it is impossible to completely define the domain of content corresponding to the target constructs and exhaustively identify all the items in the domain, domain sampling in its pure form cannot be achieved. However, there are a number of steps which can be taken to enable us to approximate domain sampling. First, since the conceptual definitions of the variables serve as a rough specification of the appropriate domains of content, they should be employed as a guide to measure development. This is what Cook and Campbell (1979, p. 64) refer to as "preoperational explication of constructs" in their discussion of construct validity, wherein they suggest that measures be tailored to fit the conceptual meaning of the target construct. Second, existing literature may be used as a source of domain content. Repeated attempts to define, theorize about, and measure a given construct gradually reveal the nature of its underlying domain of content (Bohrnstedt, 1970; Cook and Campbell, 1979; Nunnally, 1978). Given the existence of numerous published articles dealing with perceived usefulness and perceived ease of use (including those discussed in Chapter 3), prior literature represents an important source of content for measure development. To obtain elements of content from the existing literature, guided by the conceptual definitions of the target constructs, represents an approximation to domain sampling. Third, as will be discussed later, a content analysis may be performed for the purpose of improving this approximation to domain sampling.

An alternative to the present item generation approach frequently employed is to elicit items from subjects in qualitative individual or focus group interviews (e.g., Calder, 1977; Churchill, 1979; McKennell, 1974). Generating items from the literature has two of advantages over direct elicitation in the present context, however. First, there is a rich set of existing articles available to draw from, many of which have themselves employed a variety of qualitative elicitation as well as quantitative analysis techniques to understand how subjects think about these constructs. Second, these existing articles cut across a wide range of target systems, user populations and usage environments. Given the objective of creating a general model that is applicable across many contexts, the existing literature is likely to provide a more generalized representation of the desired content domains. In-depth interviews would, by necessity, be restricted to a limited user population and range of systems, which may result in highly context-specific content domains.

Thus, the measurement item candidates were generated by drawing item content from existing published studies in the Management Information Systems and Human Factors Fields. The following definitions of perceived usefulness and perceived ease of use, introduced in Chapter 2, were used as a guide for selecting which items from the literature to include in the initial pools:

Perceived Usefulness: The degree to which an individual believes that using a particular system would enhance his or her job performance.

Perceived Ease of Use: The degree to which an individual believes that using a particular system would be free of physical and mental effort.

The next step is to determine the number of items to be generated for the initial item pools. This is approached by first estimating the number of items required to achieve the desired level of reliability in the final scales, and then adding 4 additional items to account for the plan to eliminate 4 of the items based on the subsequent interviews and associated content analysis. The anticipated scale length required to achieve a Cronbach alpha reliability of .80 was estimated using the Spearman-Brown Prophecy formula:

$$\alpha = k\alpha' / [1 + (k-1)\alpha']$$

where α = desired reliability level

α' = reliability of comparable scale with n items

kn = number of items needed to achieve desired reliability

Twelve existing scales of various constructs from three published MIS studies (Ginzberg, 1981; Larcker & Lessig, 1980; Robey, 1979) were analyzed using the Spearman-Brown formula. This analysis suggested that 10 items would be required for each perceptual variable to achieve the target reliability level of .80. Thus, adding the 4-item margin needed for item elimination, it was decided to generate 14 candidate items for each variable.

The item pools for perceived usefulness and perceived ease of use are given in Tables 4.2 and Table 4.3 respectively. They are worded in terms of "electronic mail" as an example system. A wide range of published literature was drawn upon in generating the items. In addition to the empirical studies reviewed in Chapter 3, theoretical papers and reports of in-depth qualitative studies were used. Table 4A.1 in the chapter appendix specifies the articles used for abstracting the items, and Table 4A.2 gives the correspondence between these articles and specific ease of use and usefulness items.

The items within each item pool tend to have a lot of overlap in their meaning. This is expected since they are intended to be measures of the same underlying construct. Though individuals may attribute slightly different meaning to particular item statements, the goal of the multi-item approach is to

Table 4.2. Perceived Usefulness Item Pools

Item #	Item Wording
1	My job would be difficult to perform without electronic mail.
2	Using electronic mail gives me greater control over my work.
3	Using electronic mail improves my job performance.
4	The electronic mail system addresses my job-related needs.
5	Using electronic mail saves me time.
6	Electronic mail enables me to accomplish tasks more quickly.
7	Electronic mail supports critical aspects of my job.
8	Using electronic mail allows me to accomplish more work than would otherwise be possible
9	Using electronic mail reduces the time I spend on unproductive activities.
10	Using electronic mail enhances my effectiveness on the job.
11	Using electronic mail improves the quality of the work I do.
12	Using electronic mail increases my productivity.
13	Using electronic mail makes it easier to do my job.
14	Overall, I find the electronic mail system useful in my job.

downplay the effects of individual items, allowing idiosyncrasies to be cancelled out by other items, yielding a more pure indicant of the underlying construct.

Table 4.3. Perceived Ease of Use Item Pools

Item #	Item Wording
1	I often become confused when I use the electronic mail system.
2	I make errors frequently when using electronic mail.
3	Interacting with the electronic mail system is often frustrating.
4	I need to consult the user manual often when using electronic mail.
5	Interacting with the electronic mail system requires a lot of my mental effort.
6	I find it easy to recover from errors encountered while using electronic mail.
7	The electronic mail system is rigid and inflexible to interact with.
8	I find it easy to get the electronic mail system to do what I want it to do.
9	The electronic mail system often behaves in unexpected ways.
10	I find it cumbersome to use the electronic mail system.
11	My interaction with the electronic mail system is easy for me to understand.
12	It is easy for me to remember how to perform tasks using the electronic mail system.
13	The electronic mail system provides helpful guidance in performing tasks.
14	Overall, I find the electronic mail system easy to use.

Pre-test Interviews

Purpose

The purpose of the pretest interviews is to further assure content validity by empirically assessing the semantic correspondence between the measurement items contained in the item pools and the underlying variables they are intended to measure. By deriving the item pools from numerous existing studies attempting to measure the perceptual variables, we have some assurance that they provide a broad coverage spanning the domains of their respective constructs. However, we must regard this as only an approximation of what we would have obtained had we actually been able to draw sample items from their underlying content domains according to the domain sampling model.

The pretest interview's aim is to improve this approximation. Let us consider the potential deficiencies of this approximation and how it may be improved. First, although the items selected for the item pools were initially assumed to reside within the domain, it is possible that some of the items do not really belong to the domain. We can attempt to identify and remove these items by asking participants to rate the degree to which a statement corresponds in meaning to the definitions of usefulness (or ease of use) and eliminating items receiving low ratings. Recall that four additional items for each perceptual construct were added during item generation to provide for this elimination process.

Second, our selection process lacked the randomness of item selection employed by the idealized domain sampling. As a consequence, our item pools may have too much coverage in some areas of meaning, or sub-strata (Bohrnstedt, 1970, p. 92) within the domain and not enough in others. We can gather data to assess and improve the approximation to random sampling. In this case we ask subjects to rate the similarity of items to one another (using a categorization process). Based upon such data, we can infer the nature and structure of domain sub-strata, remove items in sub-strata where excess overlap exists, and add items to sub-strata where inadequate coverage is revealed.

Method

Subjects. A convenience sample of 15 subjects from the Sloan School of Management, MIT, participated in the pretest interviews. The sample included 5 secretaries, 5 graduate students and 5 members of the professional staff. All were experienced computer users.

Materials. Materials for the interviews were twenty-six 4 by 6 inch index cards. Each card had one Likert statement printed on it. The twenty-six statements corresponded to thirteen of the Likert items for each of the two perceptual variables: perceived usefulness and perceived ease of use. The fourteenth, or "overall" item for each construct was omitted since its wording was similar to the label given to the definitions of the constructs against which subjects were asked to compare the remaining items, as discussed below. Electronic mail was used as the example target system in the item wordings. A random identification number was printed on the back of each of the cards.

Procedure. The procedure was administered via face-to-face interviews and consisted of two tasks, prioritization and categorization, which were each repeated separately for the 13-card decks corresponding to usefulness and ease of use. For prioritization, subjects were first given a card upon which the label and definition of the target construct was printed and asked to read it. Next, they were asked to "rank order these 13 statements according to how well each statement's meaning matches the definition of usefulness (ease of use). Put the statement that most closely matches the meaning of usefulness (ease of use) on the top of the deck, put the statement that least matches the meaning of usefulness (ease of use) on the bottom, and so on. Electronic mail was selected as an example system only; our interest is in the meaning of the statements themselves." Extensive experience with card sorting as a data collection technique suggests that subjects find it enjoyable and exhibit high interest and concentration in the task. In the present interviews subjects appeared to find the card sorting task easy, interesting and involving to perform.

For the categorization task, subjects were asked to "put these 13 statements into categories so that items in a category are most similar in meaning to each other, and different from those in other categories. Use about 3 to 5 categories." This approach an adaptation of the "Own Categories" procedure of Sherif and Sherif (1967). Whereas Sherif and Sherif were concerned with mapping items into categories ordered along an evaluative continuum, in the present reseach we are concerned with assessing the similarity in meaning of items. That a subject places one item into the same category as another item provides a simple indicant of similarity, and requires less time and subject effort to obtain than other similarity measurement procedures such as diadic or triadic judgements.

Results and Discussion

The procedure yielded data which are summarized in 4 data matrices. Two of these contain the rankings assigned by subjects to the perceived usefulness (Table 4A.3) and perceived ease of use (Table 4A.4) items. These ranking matrices give the frequency with which the 15 subjects placed each item in a particular position in priority. The other two data matrices contain subject ratings of similarity between items for perceived usefulness (Table 4A.5) and perceived ease of use (Table 4A.6). Each cell of these symmetric matrices gives the number of subjects who put an item in the same category with some other item during the categorization task. This serves as a measure of the degree of similarity between the items as perceived by the group of subjects as a whole.

The ranking matrices (Tables 4A.3 and 4A.4) were used to derive a priority index for each item. The median rank was used as the basis for establishing priority for an item. The median was chosen in preference to the mean because of its robustness to the skewed distribution of the priority ratings. The mean was used to break ties, however. Table 4A.7 shows the medians, means and resulting priorities for perceived usefulness and perceived ease of use.

A simple cluster analysis was performed on the two similarity matrices (Tables 4A.5 and 4A.6). Items that 7 or more of the 15 subjects placed into the same category were assigned to the same cluster. For example, usefulness items 1 and 4 were coded as belonging to the same cluster. Although the derived clusters were unique in the present context, the simple clustering algorithm used may not yield unique clusters in all contexts. If not, more advanced techniques are available which do yield unique clusters (e.g., Johnson, 1967). The results of this cluster analysis are summarized in Tables 4.4 (usefulness) and 4.5 (ease of use), which gives the clusters, item numbers, item names, and item priorities. These clusters are viewed as manifestations of the underlying domain substrata, and as such serve as a basis for assessing the smoothness of domain coverage. For perceived usefulness, notice that items fall into 3 major clusters. The first cluster contains items relating to job effectiveness, the second to productivity and the third to the importance of the system to the job. If we eliminate the lowest ranked items (items 1, 4, 5 and 9), the remaining items exhibit desireable characteristics relative to the objectives of this process. Namely, important clusters (A and B), have neither too much nor too little

Table 4.4. Perceived Usefulness Item Clusters

Cluster	Item #	Item Name	Priority	Item # for Survey
A	10	Effectiveness	1	8
	3	Job Performance	2	6
	11	Quality of Work	3	1
B	12	Increase Productivity	4	5
	8	Accomplish More Work	6	7
	6	Work More Quickly	7	3
	9	Reduces Unproductive Time	10	
	5	Saves Me Time	11	
C	7	Critical to My Job	5	4
	13	Makes Job Easier	8	9
	4	Addresses My Needs	12	
	1	Job Difficult Without	13	
D	2	Control Over Work	9	2
Add	14	Overall Usefulness		10

representation of items, whereas less important clusters (C and D) do not have excess coverage.

Looking now at perceived ease of use (Table 4.5), we again find 3 major clusters. The first relates to physical effort, while the second relates to mental effort. Selecting the six highest priority items and eliminating the seventh provides solid coverage of the first two clusters. Item # 11 ("understandable"), was reworded slightly to become "clear and understandable" in an effort to pick up some of the meaning of item 1 ("confusing") which has been

Table 4.5. Perceived Ease of Use Item Clusters

Cluster	Item #	Item Name	Priority	Keep
A	8	Controllable	1	4
	10	Cumbersome	2	1
	7	Rigid & Inflexible	6	5
B	3	Frustrating	3	3
	11	Understandable	4	8
	5	Mental Effort	5	7
	1	Confusing	7	
C	12	Ease of Remembering	8	6
	4	Dependence on Manual	9	
	13	Provides Guidance	12	
D	6	Error Recovery	10	
E	9	Unexpected Behavior	11	
F	2	Error Proneness	13	
Add	14	Overall Ease of Use		10
Add		Ease of Learning		2
Add		Effort to Become Skillful		9

eliminated. The third cluster is somewhat more difficult to interpret, but appears to be tapping in to perceptions regarding how easy the system is to learn. Remembering how to perform tasks, using the manual, and relying on system guidance are phenomena which are associated with the process of learning to use a new system. Unfortunately, items 4 and 13 provide a rather indirect assessment of ease of learning. In order to correct for this, items 4 and

13 were replaced with two items that target ease of learning more directly: "ease of learning" and "effort to become skillful". Items 6, 9 and 2 were eliminated because a) they did not cluster with other items, and b) they received low priorities, and were therefore regarded as not residing within the content domain for ease of use.

In addition to the nine items remaining from this pretest interview process, an "overall" item for each construct, generated in the Item Generation process but not included in the pretest interviews, was included to provide a total of ten items per variable for the final scale.

Survey

Method

Subjects and Procedure

Subjects for the survey were 112 employees of IBM Canada's Toronto Development Laboratory. The subjects were a convenience sample of system developers, development analysts and managers. A questionnaire was circulated to 120 users on one day and collected from 112 on the following day, yielding a response rate of 93.3 %.

Questionnaire

The questionnaire (see Appendix 1) contained questions regarding two systems that are widely used in the laboratory: electronic mail and the XEDIT file editor. In order to ensure respondent familiarity with the target system, instructions in the questionnaire asked subjects not to fill out the section regarding a given system if they don't use it. Of these 112 participants, 109 completed the section of the questionnaire pertaining to electronic mail and 76 completed the section pertaining to XEDIT. For each system, respondents were asked to rate their perceived ease of use (EOU), perceived usefulness (USEF), attitude toward using (ATT) and actual current use of the system (USE).

Attitude toward using was measured using standard 7-point semantic differential rating scales as suggested by Ajzen & Fishbein (1980):

All things considered, my using electronic mail in my job is:

Neutral

Good :____:____:____:____:____:____:____: Bad

In addition, the adjective pairs Wise-Foolish, Favourable-Unfavourable, Beneficial-Harmful and Positive-Negative were used, for a total of five items making up the attitude scale. These are all adjective pairs found to load on the evaluative dimension of the semantic differential (Osgood, Suci & Tannenbaum, 1957).

Perceived ease of use and perceived usefulness were measured using the 10-item measurement scales described in the previous section. Subjects were instructed to circle the number corresponding to their responses on rating scales having the following format:

	Strongly Agree			Neutral			Strongly Disagree
I find the electronic mail system cumbersome to use.	1	2	3	4	5	6	7

Two items were used to obtain a self-reported measure of actual system use. The first one, a measure of the frequency of use of the system, had the following format:

On the average, I use electronic mail (pick most accurate answer):

Don't use at all	Use less than once each week	Use about once each week	Use several times each week	Use about once each day	Use several times each day
_____	_____	_____	_____	_____	_____

Subjects who did not use the system at all were asked to omit answers to the remaining questions pertaining to the given system (by asking them to skip to a specified page). The second usage measure asked subjects to specify about how many hours they normally spend each week using the target system. Frequency of use and amount of time spent using a target system are typical of the usage metrics routinely used in MIS research (e.g., Ginzberg, 1981; Robey, 1979). Although some existing MIS research has employed objective usage metrics from system logs, constraints of the research context did not permit the gathering of such data in the present study, restricting the study to the measurement of self-reported usage.

Results

Psychometric Properties of Scales

Reliability

The reliability coefficient used is Cronbach's coefficient alpha. Alpha was chosen over the alternative available reliability coefficients for a variety of reasons, including: (1) alpha provides a measure of the internal consistency of the items forming a multi-item scale, which is consistent with the domain sampling model by which the scales were developed; (2) it is a generalization of split-half and parallel forms coefficients; (3) compared to test-retest coefficients, it does not require 2-waves of measurement nor does it confound true fluctuations in the variable with measurement error; and (4) alpha provides a lower bound estimate of the proportion of variance in the observed measurement scale that is attributable to the variance of the true underlying construct (Bohrnstedt, 1970; Cronbach, 1951; Nunnally, 1978). As indicated previously, we set out to achieve a minimum reliability of .80 for our measurements of attitude toward using, perceived usefulness and perceived ease of use. As Table 4.6 shows, this was surpassed for attitude, usefulness and ease of use, with reliabilities generally exceeding .90. A two-item self-report scale of actual system use was also measured in the questionnaire. The hours per week question exhibited a highly right-skewed distribution of responses, and was rescaled by taking logarithms in order to make the distribution more symmetric. A linear transformation was then performed on the rescaled hours per week question to give it the same range as the frequency of use question. Averaging the use items yielded a pooled reliability of .70 and separate reliabilities of .66 and .83 for electronic mail and XEDIT respectively. The relative unreliability of the use variable has less unfavorable consequence than

Table 4.6. Cronbach Aplha Reliability of Measurement Scales

Variable	Label	# Items	Cronbach Alpha Reliability		
			E. Mail	XEDIT	Pooled
Perceived Ease of Use	EOU	10	.86	.93	.91
Perceived Usefulness	USEF	10	.97	.97	.97
Attitude Toward Using	ATT	5	.94	.97	.96
Actual System Use	USE	2	.66	.83	.70

similar unreliability would have in attitude, usefulness or ease of use since, in the structural equation modeling paradigm being employed herein, these other constructs serve as independent variables in some of the regression equations. To the extent that independent variables are measured with error, the corresponding regression coefficients tend to be biased and inconsistent (e.g., Pindyck & Rubenfeld, 1981, p. 177). In contrast, actual system use functions only as a dependent variable in the present context. Any error in measuring the dependent variable of a regression may be encompassed in the error term of the regression equation, resulting in unbiased and consistent estimates (assuming the measurement error is uncorrelated with the true disturbance, a very plausible assumption). Naturally, such error will tend to increase the error variance, although this is accounted for in the estimate of the error variance, so that hypothesis tests and confidence interval estimates remain valid (Pindyck & Rubenfeld, 1981, p. 177).

Construct Validity.

Construct validity will be evaluated using Campbell and Fiske's (1959) multitrait-multimethod technique. This technique has been widely used for the purposes of construct validation (Jaccard, Weber, & Lundmark, 1975; Ostrom, 1969; Silk, 1971). Convergent and discriminant validatation using this technique provides evidence pertinent to both content validity and common method variance. Although the multitrait-multimethod approach cannot directly evaluate the relationship between the measurement scales and the domain of content to which they are purported to correspond, it does provide useful circumstantial evidence of content validity. The failure of scales to achieve convergent and discriminant validity would cast doubt on the assumption that the scales correspond to distinct well-defined content domains. The Campbell and Fiske procedure also enables the researcher to guage the extent of method variance in the items composing scales. To the extent that an item used to measure a trait is high in method variance, it should exhibit attenuated correlation with other items of the same trait, and increased correlation with the same items applied to a different trait, which would reduce convergent and discriminant validity.

The major source of data used to assess convergent and discriminant validity is the multitrait-multimethod matrix (Campbell & Fiske, 1959), which contains the intercorrelations of the items (methods) making up a scale applied to the two different target systems, electronic mail and XEDIT (traits). For example, ease of use for one system is regarded as a distinct trait from ease of use of XEDIT. Separate multitrait-multimethod matrices were computed for each of our constructs: attitude toward using (Table 4A.8), perceived usefulness (Table 4A.9), and perceived ease of use (Table 4A.10).

Convergent validity. In order to demonstrate convergent validity, items that measure the same trait should correlate highly with one another (Campbell & Fiske, 1959). That is, the elements of the monotrait triangles (the submatrix of intercorrelations between the items intended to measure the given construct for the same system) within the multitrait-multimethod matrices should be large. The 20 monotrait-heteromethod correlations for attitude toward using were all significant, ranging from .57 to .96. Similarly for usefulness, the 90 monotrait-heteromethod correlations were all significant, ranging from .54 to .93. The monotrait-heteromethod correlations for ease of use were generally lower, falling in the range from .06 to .84, with 4 of the 90 correlations (4.4%) being nonsignificant at the .05 level (r_{12} = .14, r_{25} = .06, r_{36} = .19, r_{56} = .09). These were all for electronic mail items, which parallels our finding that the ease of use scale applied to electronic mail exhibited the lowest reliability. A likely explanation of why ease of use had some lower item correlations is because, unlike the other two motivational constructs, ease of use items were worded in both positive (e.g., "controllable") and negative (e.g., "cumbersome") directions. A separate analysis of positively and negatively worded items will be discussed later.

Discriminant validity. The multitrait-multimethod matrices (Tables 4A.8, 4A.9 & 4A.10) are also used to assess discriminant validity. The criterion is that an item should correlate more highly with other items that are intended to measure the same trait than it correlates with either the same item used to measure a different trait or different items used to measure a different trait (Campbell & Fiske, 1959). Formally, this comparison may be specified as:

$r(X_{1i}, X_{1j}), i \neq j > r(X_{1i}, X_{2k})$, where X_{1i} and X_{2i} refer to item i used to measure traits 1 and 2 respectively.

For example, within the multitrait-multimethod matrix, the correlation between items 1 and 2 measuring usefulness for electronic mail should be larger than the individual correlations between all 10 usefulness items applied to XEDIT and items 1 and 2 applied to electronic mail. For attitude toward using, the monotrait-heteromethod correlations exceeded their corresponding heterotrait-heteromethod and heterotrait-monomethod correlations for all 200 comparisons without exception. Similarly for usefulness, 1800 such comparisons were confirmed without exception. Of the 1800 comparisons for ease of use there were 58 exceptions (3%). These exception were associated with ease of use items applied to electronic mail, and involved the following items (broken down by monotrait-heteromethod vs. heterotrait-heteromethod and monotrait-heteromethod vs. heterotrait-monomethod comparisons):

Item#	MTHM vs. HTHM	MTHM vs. HTMM
1	6	0
2	10	1
4	0	1
5	16	7
6	5	1
7	4	4
9	3	0

The large number of MTHM vs. HTHM disconfirmations associated with items 1, 2 and 5 are probably due in part to the low monotrait correlations associated with these items, as discussed in the context of convergent validity above. The large number of MTHM vs HTMM exceptions for items 5 and 7 is related to the high heterotrait-monomethod correlations for these items (.33 and .30 respectively) coupled with the generally low pattern of monotrait correlations.

Table 4A.11 gives a summary histogram of the correlations comprising the multi-method matrices for attitude, usefulness and ease of use. From this table it is possible to see the clear separation in magnitude between monotrait and heterotrait correlations for attitude and usefulness, and the relatively low monotrait correlations for ease of use applied to electronic mail, resulting in an overlap with the heterotrait correlations. Also, notice that the monotrait correlations tend to be higher for XEDIT than electronic mail. This increase in convergence may have resulted from the fact that the XEDIT scales were filled out after the electronic mail scales, and the greater familiarity with the scales may have reduced random error. The histograms also show that the heterotrait-heteromethod correlations do not appear to be substantially elevated above the heterotrait-monomethod correlations. This is an additional diagnostic suggested by Campbell and Fiske (1959, p. 85) to detect the presense of method variance.

Direction-of-wording effect. The multitrait-multimethod analysis found a small proportion of exceptions to the convergent (4.4% exceptions) and discriminant (3% exceptions) validity criteria. While this would typically be strong evidence in favor of the validity of the ease of use scale (e.g., Campbell & Fiske, 1959; Silk, 1971), it is worthwhile pursuing why these exceptions occurred, and examining whether the scale can be improved. One characteristic differentiating the ease of use scale from the attitude and usefulness scales is the use of a mixture of positively and negatively worded items for the ease of use scale. The odd numbered ease of use items are framed negatively. Examination of the multitrait-multimethod matrix shows that the low monotrait correlations for ease of use for electronic mail tend to be associated with odd numbered (negative) items, and that the highest heterotrait-monomethod correlations are associated with odd numbered items (5 and 7). This suggests that convergent and discriminant validity may be improved by employing just the positive items.

Table 4A.12 gives separate histograms of the multitrait-multimethod correlations for the positive and negative ease of use items. The histograms show that the monotrait-heteromethod correlations for the positive items are higher than those for the negative items, especially for electronic mail. The magnitude of these correlations is evidence of convergent validity for the positive items, with all correlations being significant, and 9 out of 10 falling in the .50-.79 range. Moreover, it may be the case that the presence of the negative items exerted a downward influence on the correlations of the positive items. Cronbach alpha for the positive items was found to be .92 for electronic mail and .94 for XEDIT, compared to .73 and .89 respectively for the negative items. This implies greater random error for the negative items, with the error being less for XEDIT possibly due to practice effects. Two of the heterotrait-monomethod correlations were especially high (in the .30-.39 range) for the negative items, suggesting the presense of method variance. A separate assessment of discriminant validity on the positive ease of use items found no exceptions out of 200 comparisons. Cronbach aplha reliability for the positive items, pooled across systems is .92.

Another way to assess the effect of the negative items on method variance is to compare the correlations between systems on the entire ease of use scale before and after omitting the negative items. The correlation across systems was significant for the original ease of use scale ($r = .22$, $p < .05$), although not for usefulness ($r = .18$, n.s.) or attitude ($r = .09$, n.s.). The observed correlation between scores on the same scale applied to different systems may be due to a combination of "true" correlation of the underlying traits and "artifactual" correlation due to shared method variance. On theoretical grounds, we may expect there to be a "true" correlation for each of the three constructs. Ease of use may be jointly determined by the characteristics of the system as well individual characteristics such as general computer experience and intelligence. Such individual characteristics may have a similar effect on a person's perceived ease of using two different systems, producing a true trait correlation across systems. Similarly, attitude towards using computers in general may influence attitudes toward using two specific systems, and the characteristics of an individual's job may have a simultaneous influence on the perceived usefulness of two similar systems. Both of these variables may vary across subjects, producing true covariation across systems. To remove the negative ease of use items is expected to only reduce artifactual common method variance and not true variance, since the remaining scale composed of positive items has greater reliability (i.e. greater true score variance) than the original 10-item scale. When the negative items were removed from the ease of use scale, the correlation between ease of use scores across systems fell from .22 to .10. The drop in correlation is attributable to a reduction in common method variance, which was detected in the original scale by the multitrait-multimethod analysis.

Thus, using only the positive items brings the convergent and discriminant validity of the ease of use scale in line with that of the usefulness and attitude scales. The negatively worded items have a higher degree of random error and method variance. Reversing the direction of wording of items making up a scale is often advised in order to reduce the effect of methods variance (e.g. Cook & Campbell, 1979, p. 66). Ironically, just the opposite occurred here, with reversed items adding substantial method variance. Our finding parallels a finding made by Silk (1971, 393) that: "the 'reversed' item appears to be affected by method factors more than any of the other items except item 1.". Evidence suggests that direction-of-wording effects are typically much smaller than trait variance, however (Campbell, Siegman, & Rees, 1967), which is consistent with the present pattern of results. Given the substantial disadvantages of the negative ease of use items in the present context, it was decided to omit them from the ease of use scale for the analysis of the survey data.

Scale refinement. For the purposes of the experiment addressed in the following chapter, a refined ease of use scale was formed by taking the five positive items and adding a sixth positive ease of use item formed by reversing one of the existing negative items. The item which read: "The electronic mail system is rigid and inflexible to interact with" was translated into: " I find the electronic mail system to be flexible to interact with." Correspondingly, the usefulness scale was reduced in length from 10 to 6 items for the experiment. To select six of the original ten items, item analysis was performed by examining corrected item-total correlations which were calculated by removing the item from the rest of the items in the scale prior to computing the correlation. The items having the five highest correlations were selected and combined with the "overall usefulness" item (# 10) for the final usefulness scale. These were items 3, 5, 6, 8, and 9 (with corrected item-total correlations of 87.5, 93.0, 93.0, 93.0, and 87.0 respectively). The Spearman-Brown formula estimates that this should yield a reliability of .94 for the revised usefulness scale. In order to ensure that the reduced scales still represented the appropriate domains of content, the clusters corresponding to the selected items, found via pretesting above, were examined. The 6-item ease of use scale contained 2 items associated with cluster A, 1 with cluster B, and 2 with cluster C. Neither of the 2 non-included B items could have been converted easily from negative to positive in wording. For the usefulness scale, 2 items corresponded to cluster A, 2 to cluster B and 1 to cluster C. For both scales, the "overall" item was included, but was not part of the cluster analysis. Thus, the revised scales continue to span the inferred content domain substrata identified in the pretest.

Summary. The multitrait-multimethod analysis found very high levels of convergent and discriminant validity for the scales used in the present research. After eliminating the negatively worded ease of use items, there were no exceptions found to the convergent and discriminant validity criteria. All monotrait correlations were significant and high, and were all greater in value than their corresponding heterotrait correlations. This is an unusally high level of validity, and many scales are considered to be quite healthy despite minor departures from the criteria (e.g., Campbell & Fiske, 1959; Silk, 1971). We regard this as evidence that the scales are not materially invalidated by method variance, and as circumstantial evidence of content validity of the scales. In addition, the scales exhibited reliabilities in excess of .90.

Empirical Test of Technology Acceptance Model

Regression analyses were performed on data pooled across the two target systems (n = 185) to test the hypothesized relationships of the model. Table 4.7 contains the results of OLS regressions applied to the hypothesized equations of the model, and Table 4.8 contains the unrestricted regressions needed to carry out the hierarchical regression test of the insignificance of those causal relationships hypothesized to be nonsignificant. Several of the hypotheses

Table 4.7. Survey Data- TAM Regression Tests

Dep. Var.	R^2	Independent Variable	b	S.E. (b)	β	t. stat.	sig. lvl.
EOU	.044	Constant	3.323	.155		21.463	.000
		System	-.581	.201	-.210	-.289	.004
USEF	.400	Constant	.933	.214		4.356	.000
		System	-.036	.151	-.014	-.238	.812
		EOU	.591	.055	.630	10.661	.000
ATT	.550	Constant	.224	.134		1.668	.097
		EOU	.100	.049	.134	2.037	.043
		USEF	.531	.054	.650	9.893	.000
USE	.308	Constant	4.192	.283		14.802	.000
		ATT	.220	.025	.555	8.792	.000
USE	.361	Constant	3.411	.323		10.565	.000
(w/ USEF		USEF	.077	.016	.435	4.913	.000
included)		ATT	.089	.039	.205	2.316	.022

were confirmed by the data: system had a significant effect on ease of use ($t = .289$, $p < .01$); ease of use had a significant effect on both usefulness ($t = 10.66$, $p < .01$) and attitude ($t = 2.04$, $p < .05$); usefulness had a significant effect on attitude ($t = 9.89$, $p < .01$); and attitude had a direct effect on usage

Table 4.8. Survey Data- Regression Tests of hypothesized Insignificant Relationships for TAM

Dep. Var.	R^2	Independent Variable	b	S.E. (b)	β	t. stat.	sig. lvl.
ATT	.574	Constant	.484	.155		3.121	.002
		System	-.323	.103	-.159	-3.133	.002
		EOU	.077	.049	.104	1.599	.112
		USEF	.532	.052	.651	10.155	.000
USE	.365	Constant	3.010	.416		7.235	.000
		System	.366	.278	.084	1.314	.191
		EOU	.147	.130	.092	1.129	.261
		USEF	.669	.175	.380	3.829	.000
		ATT	.449	.201	.206	2.233	.027

behavior ($t = 8.79$, $p < .01$). Also consistent with expectation, the system-use and eou-use relationships were found to be non-significant.

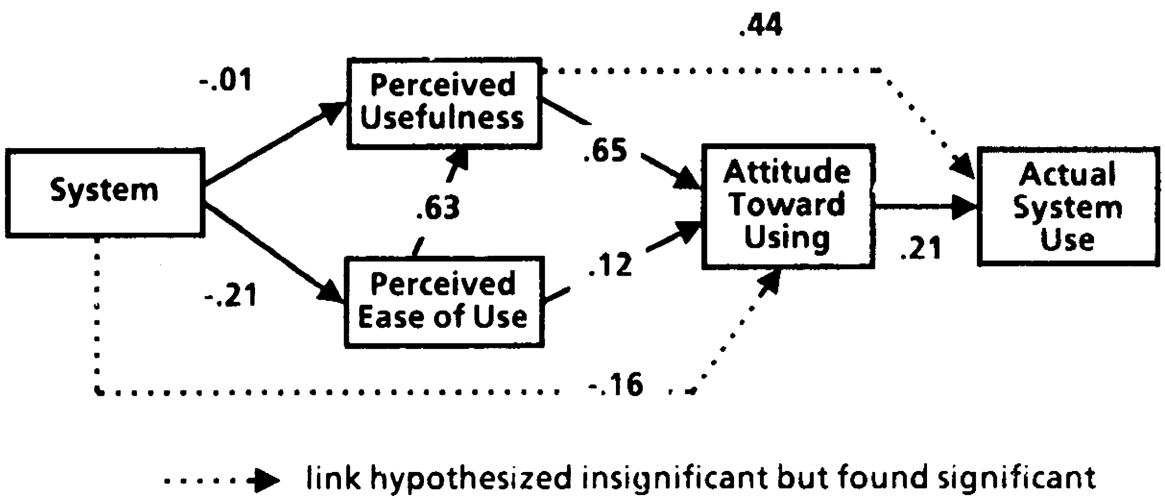
Contrary to expectation, system exerts a direct influence on ATT over and above USEF and EOU ($t(181) = 3.13$, $p < .01$). Also counter to expectation, the hypothesis that no variables other than attitude influence was rejected using hierarchical regression ($F(3, 180) = 5.39$, $p < .01$). From Table 4.8, there appears to be a significant direct effect from USEF to use ($t(180) = 3.83$, $p < .01$). Therefore, USEF was included in the regression for estimating the attitude-use relationship in order to provide a consistent estimate (Table 4.7). The theoretical interpretation of a direct USEF-USE link will be dealt with in the discussion section. As hypothesized, no direct effect of System or EOU on USE was observed ($F(2,180) = .00$, n.s.). Table 4.9 gives the point estimates and confidence intervals based on these regressions. The parameters are specified in the causal diagram of Figure 4.1. The parameters enable one to compute the

Table 4.9. TAM Parameter Estimates and 95% Confidence Intervals- Survey Data

Causal Link		Point Estimate			95% Confidence Interval	
Ind. Var.	Dep. Var.	β	Std. Error	Sig. Level	Lower Bound	Upper Bound
System	EOU	-.210	.073	.004	-.352	-.068
System	USEF	-.014	.059	.812	-.129	.101
EOU	USEF	.630	.059	.000	.515	.745
System	ATT	-.159	.051	.002	-.259	-.059
EOU	ATT	.104	.066	.112	-.027	.235
USEF	ATT	.651	.064	.000	.524	.778
USEF	USE	.435	.090	.000	.258	.612
ATT	USE	.205	.090	.022	.029	.381

relative importance of USEF and EOU in influencing USE. USEF has both a direct effect (.44) plus an indirect effect via ATT (.65 x .21). Combined, this equals .58. EOU has an effect on USE through ATT: .12 x .21; plus an effect through USEF: .63 x .58 (.58 from above calculations of USEF's effect on USE). This totals .39. Comparatively, therefore, USEF is 1.49 times as important as EOU in influencing USE in the survey data.

Figure 4.1. Causal Diagram of Model Validation Results- Survey



Discussion

The model testing analysis confirmed several of the hypotheses and disconfirmed others. The TAM motivational variables: attitude toward using, perceived usefulness and perceived ease of use, taken together, fully mediate between system design features and self-reported usage behavior. That is, the characteristics of the system appear to influence behavior entirely through these motivational variables and have no additional direct effect on use. From the standpoint of the overall research objectives of identifying the mechanisms linking design features to usage, the model has been successful. However, the specific pattern of relationships found between the individual variables deviated from expectations in some cases. First, although usefulness was expected to influence system use only indirectly through its effect on attitude, the data was inconsistent with this hypothesis, suggesting instead that usefulness has a significant direct effect on use in addition to its indirect effect via attitude. Second, the hypothesized significant effect of system on usefulness was found to be nonsignificant. Third, the influence of ease of use on attitude was hypothesized to be significant but found nonsignificant. Fourth, whereas perceived usefulness and perceived ease of use were hypothesized to fully mediate between system characteristics and attitude toward using, the data suggests that the characteristics of the system have a direct effect on attitude. A theoretical interpretation of this pattern of findings follows.

The direct effect of a perception on behavior over and above its indirect effect through attitude, such as the observed usefulness-use link, is inconsistent with the specification of the Fishbein model (e.g., Fishbein & Ajzen, 1975, p. 314). While considerable evidence has accumulated in support of the argument that beliefs (and corresponding evaluation weights) determine attitude, that attitude (along with social norm) determines behavioral intention, and that intention determines overt behavior (e.g., Davidson & Morrison, 1983; Fishbein, 1966; Ajzen & Fishbein, 1970, 1972, 1973, 1980; McArdle, 1972; Jaccard & Davidson, 1972; Warshaw, 1980a; 1980b; Warshaw & Davis, 1985), there has been comparatively little attention directed to the question of whether beliefs may have a direct effect on intentions and/or behavior. A review of the pertinent literature revealed some work on the subject, however. An alternative to the Fishbein model, specified by Triandis (1977, p. 194), views cognitions as having a direct effect on behavioral intentions:

$$BI = w_s S + w_a A + w_c C$$

where S = social factors

A = affect attached to the behavior, and

C = perceived consequences of the behavior

Note that Triandis's formulation does not model any direct effect from cognition to affect as specified by the Fishbein model. Two studies comparing the Fishbein and Triandis formulations found mixed evidence regarding the direct influence of cognitions on intention. Jaccard and Davidson (1975) compared the Triandis and Fishbein models in the context of family planning behavior, and found that the addition of cognitions (perceived consequences) did not account for significant additional explained variance in behavioral intention. Brinberg (1979) performed a similar comparison regarding church attendance behavior, finding that perceived consequences entered into the regression equation at the .01 level of significance for Protestants, although were nonsignificant for Jewish and Catholic respondents. Triandis (1977, p. 197) argues that the weight of the cognitive component should be influenced by the extent to which the target behavior is connected with consequences having large positive or negative value. The usefulness variable in the present research models the effect of computer usage behavior on gains in work performance. In most organizational settings, reward systems have significant reinforcing (and punishing) capability, and are linked to work performance (Vroom, 1964). From this perspective, the fact that the usefulness variable in the present research models the effect of computer usage behavior on gains in work performance would tend to increase the expected magnitude of the associated direct cognition-intention relationship. Given that intention is not included in the present model (due to reasons discussed in Chapter 2), the expected direct cognition-intention link translates into a direct cognition-behavior link.

Bagozzi (1982) specified a hybrid intention model which combines aspects of the Triandis and Fishbein perspectives by representing both a direct effect of cognitions on intentions and an indirect effect via affect. In his blood donation study, he found a significant cognition-intention link, controlling for affect. He theorizes that: "The direct path from expectancy-value judgements to intentions is hypothesized to operate through stored imperatives in memory. One possibility is that the expectancy-value judgements activate a personal goal, and the goal, in turn, influences one's intention to act. (p.575)" Such goal activation may be operative in the present context in that organizational respondents would be expected to have well-defined cognitive representations of the consequences of increasing their work performance (e.g., nature of rewards, affect toward those rewards (Vroom, 1964)). Subjects may cognitively process the opportunity to use a new computer system as an instrumental act that may lead them toward a goal for which they expect to receive valued rewards. Such a cognitive appraisal may invoke the corresponding stored goal in the subject's memory without fully arousing the affect associated with the rewards of attaining the goal. Other potential mechanisms for a direct cognition-intention relationship mentioned by Bagozzi (1982) are values, scripts, social norms, and novelty-seeking urges.

Thus, available theory and data suggest that a direct link between cognitions and behavior may exist under certain circumstances, notwithstanding the original Fishbein model specification. Moreover, the perceived usefulness variable appears to fit the characteristics associated with such effects. Coupled with the strong direct effect observed in the data, this suggests that the TAM specification should be revised to incorporate the direct usefulness-use link as a pemanent feature of the model.

The fact that usefulness exerts more that twice as much direct influence on use than does attitude toward using (with regression coefficients of .44 and .21 for usefulness and attitude respectively) underscores the importance of the usefulness variable. In addition, usefulness exerts 3 times as much influence on attitude as does ease of use. In fact, ease of use does not have a significant direct effect on attitude, as hypothesized, instead influencing attitude only indirectly via its relatively strong effect on usefulness (.64). This gives cause to rethink the role of the usefulness variable. Referring back to the definition of usefulness as "The degree to which an individual believes that using a particular system would enhance his or her job performance", it seems that usefulness may be regarded as a "net" construct which reflects considerations of the "benefits" as well as the "costs" of using the target system (e.g., Einhorn & Hogarth, 1981; Johnson & Payne, 1985). Ease of use (or more appropriately its inverse, effort of using) may be seen as part of the cost of using the system from the subject's perspective. This viewpoint would account for why ease of use operates entirely through usefulness in the survey data. In the experiment described in the next chapter, alternative models specified according to this "net cost-benefit" view of the usefulness variable will be addressed.

The lack of a significant effect of system on may have be a reflection of the fact that the two test systems used simply did not provide a strong enough manipulation of usefulness. Although the systems clearly differed in terms of their functional characteristics, one being a mail system, the other a file editor, they appear to have greater similarity in terms of usefulness. Perhaps this should not be too surprising, however. Both systems have been"accepted" by the user subjects, and hence have been "pre-selected" from among a larger set of potential systems, many of which probably had inferior usefulness. The fact that these systems have been accepted may have biased the selection of systems in such a way that their usefulness levels are more homogeneous than might be found in an applied user acceptance testing situation, where both potentially successful and unsuccessful systems are being analyzed in an effort to predict which ones will more readily achieve acceptance.

The significant direct influence of system characteristics on attitude toward using implies that perceived usefulness and perceived ease of use alone may provide an incomplete account of the cognitive mechanisms mediating between system and attitude. This leads us to consider possible variables that should be added to the model. The previous discussion has emphasized the importance of perceived usefulness, arguing that ease of use operates through this variable. Thus the model views computer usage behavior to be largely instrumental, being driven by concern over performance gains and associated rewards. Computer usage is therefore being treated as an extrinsically motivated behavior. Recently, Malone (1981) has drawn attention to the fact that intrinsic motives play an important role in determining usage of at least some kinds of computer systems. That is, people use systems in part because they enjoy the process using them (and thereby gain intrinsic reward), not just because they are being extrinsically rewarded for the consequences of usage behavior. Intrinsic motives may be the mechanism underlying the observed direct effect of system characteristics on attitude toward using. From this perspective, an individual's affect toward using a given system would be expected to be jointly determined by the extrinsic and intrinsic rewards of using the system. Malone (1981) discusses design characteristics of systems which are expected to influence intrinsic motivation. The experiment in the next chapter examines alternative models which incorporate a variable called "anticipated enjoyment of using" in order to address the issue of instrinsic motivation.

Chapter 4 Appendix. Tables.

Table4A.1. Articles Used for Usefulness and Ease of Use Item Generation

#	Author (s) & Year	#	Author (s) & Year
1	Bailey & Pearson, 1983	21	Lucas, 1978
2	Barrett, et al., 1968	22	Magers, 1983
3	Bewley, et al., 1983	23	Maish, 1979
4	Brosey & Shneiderman, 1978	24	Malone, 1981
5	Butler, 1983	25	Mantei & Haskell, 1983
6	Card, et al., 1980	26	Miller, 1977
7	Carroll & Carrithers, 1984	27	Neal & Simons, 1984
8	DeSanctis, 1983	28	Poller & Garter, 1983
9	Dzida, et al., 1978	29	Poppel, 1982
10	Fuerst & Cheney, 1982	30	Rice, 1980
11	Gallagher, 1974	31	Roberts & Moran, 1983
12	Ginzberg, 1981	32	Rossen, 1983
13	Good, 1982	33	Schewe, 1976
14	Gould & Boies, 1984	34	Smith, et al., 1982
15	Gould & Lewis, 1985	35	Schultz & Slevin, 1975
16	Guthrie, 1973	36	Swanson, 1974
17	Kaiser & Srinivasan, 1982	37	Zmud, 1978
18	Keen, 1981		
19	King & Epstein, 1983		
20	Larcker & Lessig, 1980		

Table 4A.2. Correspondence Between Usefulness and Ease of Use Scale Items and Articles from Which They Were Generated

Item #	Article # from Table 4A.1.	
	Usefulness	Ease of Use
1	11, 20	9, 11, 22
2	23, 30, 35	3, 4, 6, 7, 9
3	1, 8, 12, 35	7, 22, 27
4	16, 17, 35	1, 9, 22
5	18, 29, 30	23, 32
6	18, 30	1, 9, 22, 23
7	11, 20, 35	1, 9, 22, 23
8	14, 18, 29	1, 9, 22
9	21, 35	9, 11, 24, 25
10	10, 23	6, 31
11	8, 21, 18	3, 11, 19, 36
12	23	3, 9, 28
13	35	9, 22, 34
14	1, 11, 15, 23, 35, 37	1, 2, 15, 20, 23, 26

Table 4A.3 Ranking of Item Meaning for Perceived Usefulness: Frequency by Item

Item #	Ranked Correspondence with Construct Meaning (1 = highest)												
	1	2	3	4	5	6	7	8	9	10	11	12	13
1	1			2		2			1	1	1	3	4
2	1		1			2	2	3		3	2		1
3	4	1	4	1	1				1	1		2	
4		1	1	2	1		1		1		1	5	2
5	1			1	1	1	1	1	2	3	2	2	
6				3	2	3	2	2			1	1	1
7	3	2		1	2		1	1		1	2		2
8			2	1	1	2	3	4			2		
9						2	1	1	5		2		4
10	2	3	4	1		1		1	1	1	1		
11	2	4	1	1	3		1		1		1	1	
12	2	1		1	2	1	3	1	1	2	1		
13		3	1	1	2	1		1	2	2		1	1

Table 4A.4 Ranking of Item Meaning for Perceived Ease of Use: Frequency by Item

Item #	Ranked Correspondence with Construct Meaning (1 = highest)												
	1	2	3	4	5	6	7	8	9	10	11	12	13
1			1	1	1	2	5	1		1		3	
2				1			1	3	3	1	3	2	1
3	2	1	2	3	2				1	1	1		2
4	2				2	2	1	1	1	2	1	2	1
5	1	2	1		2	2	2				4		1
6			2	1	1	2		1	2	2	1	1	2
7		2	2	2	1	1	1	1		1	1	1	2
8	6	2	1	2	1		1			1		1	
9				1	1	2	1	3	1	2		3	1
10	1	2	2	2	3	1	1	2		1			
11	2	2	1	1	1	3		1	1				3
12		4	1		1	1	1		3	2	2		
13	1		2	1			1	1	2	2	2	2	1

Table 4A.5. Similarity Matrix for Perceived Usefulness Items: Frequency With Which Items Were Assigned to Same Category

ITEM	1	2	3	4	5	6	7	8	9	10	11	12	13
1													
2	1												
3	3	3											
4	7	5	1										
5	2	0	0	0									
6	1	0	0	0	13								
7	7	4	2	11	0	0							
8	1	1	2	2	7	9	0						
9	1	1	2	0	9	11	0	8					
10	1	2	11	0	1	1	2	1	0				
11	1	3	11	0	0	0	2	0	1	11			
12	0	0	3	0	6	8	0	11	7	3	2		
13	7	4	2	2	3	2	2	2	1	4	2	1	

Table 4A.6 Similarity Matrix of Perceived Ease of Use Items: Frequency With Which Items Were Assigned to Same Category

ITEM	1	2	3	4	5	6	7	8	9	10	11	12	13
1													
2	3												
3	7	3											
4	3	3	2										
5	**9**	3	5	2									
6	0	6	1	1	1								
7	2	2	5	1	2	1							
8	2	1	2	1	1	2	7						
9	2	5	3	2	1	2	4	2					
10	3	2	4	0	4	2	**9**	5	5				
11	**8**	1	4	2	6	3	1	6	3	2			
12	1	1	0	**8**	4	3	1	4	1	1	6		
13	4	1	1	**8**	3	1	0	2	1	2	3	6	

Table 4A.7 Determination of Item Priorities for Perceived Usefulness and Ease of Use

ITEM	PERCEIVED USEFULNESS			PERCEIVED EASE OF USE		
	Median Rank	Mean Rank	Priority	Median Rank	Mean Rank	Priority
1	12	9.3	13	7	7.5	7
2	8	7.9	9	10	9.5	13
3	3	4.7	2	5	5.9	3
4	12	8.7	12	9	7.7	9
5	10	8.3	11	7	6.7	5
6	7	7.1	7	9	8.1	10
7	7	6.3	5	7	6.9	6
8	7	6.8	6	3	3.7	1
9	9	9.7	10	9	8.7	11
10	3	4.5	1	5	4.9	2
11	5	4.7	3	6	6.1	4
12	7	5.7	4	9	6.5	8
13	8	6.7	8	10	8.2	12

Table 4A.8. Multitrait-multimethod Matrix of Item Intercorrelations - Attitude Toward Using

ITEM	Attitude Toward Using-Electronic Mail (MAU) 1	2	3	4	5	Attitude Toward Using - Xedit (XAU) 1	2	3	4	5
MAU1										
MAU2	.72									
MAU3	.70	.72								
MAU4	.62	.75	.75							
MAU5	.57	.78	.71	.82						
XAU1	-.10	-.04	.14	-.01	.11					
XAU2	-.04	-.01	.15	.03	.13	.85				
XAU3	-.01	.02	.15	.02	.13	.84	.95			
XAU4	.02	.05	.18	.08	.16	.80	.92	.94		
XAU5	.00	.02	.21	.04	.15	.84	.94	.95	.96	

Table 4A.9 Multitrait-multimethod Matrix of Item Intercorrelations- Perceived Usefulness

ITEM	Perceived Usefulness- Electronic Mail (MUF)										Perceived Usefulness- Xedit (XUF)									
	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
MUF1																				
MUF2	.77																			
MUF3	.68	.60																		
MUF4	.57	.59	.54																	
MUF5	.70	.62	.80	.67																
MUF6	.69	.77	.69	.65	.76															
MUF7	.68	.65	.77	.67	.82	.76														
MUF8	.73	.73	.76	.76	.81	.79	.87													
MUF9	.77	.71	.75	.72	.79	.76	.85	.87												
MUF10	.65	.58	.68	.62	.80	.62	.71	.77	.74											
XUF1	.24	.31	.26	.25	.18	.34	.34	.38	.38	.23										
XUF2	.18	.27	.15	.21	.12	.32	.26	.33	.29	.20	.85									
XUF3	.09	.17	.05	.22	.00	.20	.15	.23	.20	.11	.85	.90								
XUF4	-.07	.00	-.03	.24	-.05	.13	.12	.13	.07	.01	.68	.65	.73							
XUF5	.08	.20	.12	.23	.07	.27	.24	.29	.26	.17	.85	.86	.90	.80						
XUF6	.07	.18	.05	.27	.04	.25	.23	.29	.21	.12	.82	.84	.86	.80	.92					
XUF7	-.07	.05	.00	.19	-.04	.12	.11	.17	.11	.00	.67	.71	.80	.74	.85	.86				
XUF8	.08	.16	.04	.23	.03	.24	.22	.27	.22	.14	.77	.83	.86	.75	.89	.93	.85			
XUF9	.04	.13	.03	.01	-.03	.05	.02	.09	.09	.13	.73	.80	.83	.60	.82	.79	.76	.80		
XUF10	.00	.10	.05	.08	.00	.16	.17	.16	.16	.13	.80	.79	.83	76	.88	.86	.79	.83	.86	

Table 4A.10. Multitrait-multimethod Matrix of Item Intercorrelations - Perceived Ease of Use

ITEM	Perceived Ease of Use - Electronic Mail (MEU)										Perceived Ease of Use - Xedit (XEU)									
	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
MEU1																				
MEU2	.14																			
MEU3	.35	.23																		
MEU4	.43	.51	.25																	
MEU5	.38	.06	.49	.29																
MEU6	.26	.47	.19	.57	.09															
MEU7	.24	.22	.43	.30	.32	.27														
MEU8	.41	.57	.35	.77	.29	.68	.34													
MEU9	.37	.32	.42	.30	.30	.30	.48	.35												
MEU10	.40	.61	.32	.75	.25	.62	.35	.78	.47											
XEU1	.19	.12	.12	.21	.11	.15	.19	.16	.09	.21										
XEU2	.08	.19	.06	.15	.11	.02	.13	.01	.12	-.03	.51									
XEU3	.10	.11	.17	.20	.11	.07	.25	.15	.18	.20	.73	.40								
XEU4	.24	.10	.05	.29	.01	.14	.14	.18	.07	.14	.78	.62	.67							
XEU5	.15	.06	.13	.18	.33	.04	.27	.04	.31	.10	.63	.46	.63	.55						
XEU6	.06	.20	.11	.11	.09	.10	.04	.05	.08	.01	.57	.69	.44	.67	.42					
XEU7	.23	.10	.16	.22	.15	.10	.30	.15	.17	.15	.65	.37	.68	.59	.72	.42				
XEU8	.10	.11	.06	.09	.10	-.01	.06	-.04	.07	-.04	.64	.78	.52	.73	.54	.82	.46			
XEU9	.09	.02	.09	.09	-.10	.20	.07	.07	.15	.09	.63	.37	.57	.52	.52	.44	.64	.42		
XEU10	.14	.12	.07	.06	[illegible]	-.05	.05	-.05	.11	.01	.70	.68	.66	.77	.58	.78	.50	.84	.49	

Table 4A.11. Multitrait-multimethod Correlations by Construct, Type and Size:

Correlation Size	Construct											
	Attitude Toward Using				Perceived Usefulness				Perceived Ease of Use			
	Same Trait/ Diff. Meth.		Different Trait		Same Trait/ Diff. Meth.		Different Trait		Same Trait/ Diff. Method		Different Trait	
	Elec. Mail	Xedit	Same Meth.	Diff Meth.	Elec. Mail	Xedit	Same Meth.	Diff. Meth.	Elec. Mail	Xedit	Same Meth.	Diff. Meth.
-.20 to -.11												1
-.10 to -.01			2	4				6			1	5
.00 to .09			1	8			3	25	2		1	32
.10 to .19			2	7			2	27	2		5	40
.20 to .29				1			5	25	9		1	11
.30 to .39								7	14	2	2	1
.40 to .49									9	9		
.50 to .59	1				4				3	11		
.60 to .69	1				14	4			3	13		
.70 to .79	7				20	11			3	8		
.80 to .89	1	4			7	26				2		
.90 to .99		6				4						
# Correlat'ns	10	10	5	20	45	45	10	90	45	45	10	90

Table 4A.12. Multitrait-multimethod Correlations for Positive and Negative Ease of Use Items

Correlation Size	Positive Items				Negative Items			
	Monotrait-heteromethod		Heterotrait		Monotrait-heteromethod		Heterotrait	
	E-mail	XEDIT	Mono method	Hetero method	E-mail	XEDIT	Mono method	Hetero method
-.20 to -.11								
-.10 to -.01			1	5				1
.00 to .09			1	6				4
.10 to .19			2	8			3	11
.20 to .29			1	1	1			4
.30 to .39					5		2	
.40 to .49	1				4			
.50 to .59	3					2		
.60 to .69	3	4				6		
.70 to .79	3	4				2		
.80 to .89		2						
.90 to .99								
# Correlat'ns	10	10	5	20	10	10	5	20

CHAPTER 5. EXPERIMENT

Objectives

There are four major objectives for this experiment: (A) to validate videotape presentation as a viable substitute for hands-on interaction as a medium for demonstrating systems in a user acceptance testing context; (B) to further validate the causal structure of the proposed technology acceptance model (TAM); (C) to evaluate and compare the specific systems employed in the experiment, and (D) to explore a series of theoretical elaborations and extensions to the existing technology acceptance model. These objectives are discussed below.

Comparison of Videotape Presentation with Hands-on Interaction

TAM was primarily designed to be employed in user acceptance testing experiments in which subjects receive a hands-on demonstration of the prototype systems and then rate their reactions. The purpose of the hands-on demonstration is to provide subjects a realistic, behaviorally-based exposure to the system from which they can form stable attitudes and perceptions (Fazio & Zanna, 1978, 1981). The purpose of the video validation component of the experiment is to address the feasibility of using videotapes in place of hands-on demonstrations as a medium for presenting prototype systems to potential users.

Videotapes have a number of attractive advantages relative to hands-on demonstration of prototype systems for user acceptance testing:

1) Videotapes enable testers to present hypothetical systems which may not physically exist, by simulating the system interface. The outcome of a video-based acceptance test can then be used as an input to

decisions regarding which of the hypothetical system capabilities should be built.

2) In cases where prototypes actually do exist, they may not be in enough of a "final product" form to obtain untainted user judgements. For example, if software precautions necessary to trap certain classes of user errors are planned, but have not yet been implemented, experimental participants may inadvertantly get stuck in undesired states from which they cannot proceed toward completing their task without external help. This could cause negative perceptions regarding ease of use which may not be reflective of the final product. (Conversely, however, videotapes may unknowingly disguise interface flaws that go unnoticed by designers, and do not get conveyed to subjects viewing a video demonstration).

3) Videotapes are more portable, making it easier to run acceptance tests at remote sites.

4) Videotapes make it more feasible to run larger numbers of subjects in parallel, being constrained not by the number of test systems, but rather by the number of subjects that can view a projection screen.

5) Less time appears to be required to cover the same amount of information by video than by live demonstration, which reduces the testing time per subject.

6) Due to the reduced time requirements, more system versions can be shown to each subject, thereby generating more useful information for the same number of subjects.

Given the numerous advantages of videotape relative to hands-on demonstration, it is desireable to assess the degree to which it can serve as a substitute. A major potential disadvantage is that the passive viewing of a system and it's operation via videotape may be less realistic that the active hands-on use of the same system, resulting in less accurate prediction of ultimate user adoption of the system. Psychological research has frequently observed that attitudes based on indirect experience with the attitude object are less stable, less well-formed and less predictive of future behavior than are attitudes based on direct behavioral experience with the attitude object (Fazio & Zanna, 1981). Thus, attitudes and perceptions measured from a subject receiving a video demonstration of the target system can be viewed as an approximation of what that same subject's attitudes and perceptions would have been had he or she received a behavioral demonstration. The key research question in validating video methodology is therefore: "how good is this approximation?" We expect this approximation to be different for usefulness and ease of use. That is, we expect it to be more feasible to convey by videotape the nature of what the system is intended to perform in such a manner that the subject can judge it's applicability to his or her job (i.e. usefulness) than to convey the amount of effort required to actually interact with the system (i.e. ease of use). In addition, since perceived usefulness is hypothesized to have greater influence on attitudes that does ease of use (based on findings from the prior survey), we expect videotape based attitude and behavioral expectation ratings to be good approximations of those based on hands-on interaction. This can be expressed in the following hypothesis:

H1: Presentation medium (video vs. hands-on) will have a significant effect on ease of use ratings, but not on usefulness, attitude or behavioral expectation ratings.

Thus, the videotape validation will obtain a statistical estimate of the degree to which videotape demonstration approximates hands-on demonstration for the systems being studied. This information is intended to provide a basis for addressing whether video demonstrations have a viable role in user acceptance testing, and for analyzing the accuracy and testing efficiency tradeoffs in comparison to hands-on demonstration.

As part of the assessment of video's ability to substitute for hands-on, subjects' confidence in their ratings will be analyzed. Prior research (e.g., Fazio & Zanna, 1978; 1981) has established that when perceptions and attitudes are more stable and well-formed, as when they are based on direct experience, they are held with greater confidence. Researchers have repeatedly found that more confidently held attitudes are more stable and more predictive of future behavior. In the present context, if we can establish that attitude confidence enable us to tell how well video based ratings will predict hands-on based ratings, the confidence variable may serve as an important diagnostic in future contexts where only video-based measurements are taken. Low confidence ratings would caution us about the accuracy with which we may infer direct-experience based perceptions and attitudes from video based ratings. An important first step, therefore, is to verify that attitude confidence functions as an indicant of stable, well-formed attitudes in the user acceptance testing context. Operationally, we will examine whether confidence in one's video ratings moderates the prediction of hands-on ratings from video ratings.

Further Validation of Technology Acceptance Model (TAM)

Validation of any theoretical model entails taking reliable measurements of the model's variables and assessing the degree to which the observed statistical relationships between the variables are consistent with the hypothesized causal structure of the model. Reliable measures of the TAM motivational variables: perceived ease of use, perceived usefulness, and attitude toward using, have already been developed and reported upon in Chapter 4. In addition, this experiment employs a new variable that has not previously been included in the Technology Acceptance Model, called behavioral expectation (BE). This new variable was recently established in the literature by Warshaw & Davis (1985; in press), and is considered to be the best single predictor of one's future behavior. In essence, it is one's own self-prediction of what one will do in the future. In the TAM context, it will be used to measure self-predicted use of the target system. Due to the inability to measure actual organizational acceptance of the systems in the experiment, BE will be included in the model as a substitute for actual behavior.

The present experiment generated additional evidence regarding the validity of the TAM relationships. The experimental data can help determine whether or not the absense of a direct observed effect from ease of use to attitude in the pre-test survey was a methodological artifact. One potential artifact is that, due to the need for prior subject exposure to the target systems, the subjects for the pre-test survey were existing users (and therefore acceptors) of the target systems. Hence individuals who had considered using but ultimately rejected the systems may have been excluded. It is possible that some of the latter rejected the system due to difficulty of use, and that if they had been included in the sample we may have observed a greater influence of ease of use on attitude. If so, then we may find that a strong direct ease of use --> attitude relationship appears in the experimental data, although it was not found in the pre-test survey. Since TAM is intended to be a descriptive model of user motivation in the context of user acceptance testing experiments, it should encompass both potential acceptors and rejectors of the target technology. The present experiment will employ as subjects both potential acceptors and rejectors, and thus provides an empirical context for model validation that more closely matches the intended context of TAM.

Regarding the direct usefulness-use link observed in the survey, although this is counter to the Fishbein and Ajzen (1975) paradigm, a further review of related theoretical literature provided an alternative account of mechanisms through which a direct cognition-behavior effect could occur (Bagozzi 1982; Triandis, 1977; see Chapter 4). This analysis, coupled with the strong effect observed in the survey, suggests that a direct link from usefulness to behavioral expectation will be observed again in this experiment. Thus, the following hypotheses are tested:

H2: Attitude toward using and perceived usefulness will both have a significant effect on behavioral expectation.

H3: Ease of use will not have a significant effect on behavioral expectation over and above the effect of attitude.

H4: Usefulness and ease of use will both have a significant effect on attitude toward using.

H5: Ease of use will have a significant effect on usefulness.

H6: The system will have a significant direct effect on usefulness and ease of use, although it will not have a significant direct effect on attitude or behavioral expectation (controlling for indirect effects via usefulness and ease of use).

Evaluation of Specific Systems Tested

Although the major purposes of the experiment are to validate videotape presentation and to validate TAM, the target systems to be used for the experiment were selected in part due to interest in comparing them against each other per se. Although we do not hypothesize which system will fare better, because of the discriminant validity of the measures to be used (Davis, 1985), significant differences between systems are expected.

Pendraw

Pendraw is an IBM PC based software package made by Pencept, Inc. of Waltham Massachusetts, that enables users to draw a wide variety of different types of visual aids. The package uses the Pencept Penpad digitizer tablet as its primary interface device. The digitizer tablet accepts users penstrokes as inputs, is capable of performing character recognition for inputing characters, and has a command template located near the top of its surface for invoking functions. The user manipulates the image which is visible on the screen as it is being created. Pendraw enables the user to create drawings using freehand drafting as well as geometric shapes such as boxes, lines and circles. Features for copying, moving and erasing sections of the image are provided. A variety of line widths, color selections and title fonts are available for the user to select from. In addition, Pendraw enables the user to capture and manipulate images created using other programs.

Chartmaster

Chartmaster is an IBM PC based software package made by Decision Resources, Inc. of Westport, Connecticut, that enables users to create numerical charts such as bar charts, line charts and pie charts. The keyboard is used as the input device. The system guides the user through a series of menues by which the user inputs the data for, and defines the desired characteristics of, the chart to be made. The chart can then be plotted on the screen. The user can specify a wide variety of options relating to title fonts, colors, plot orientation, crosshatching pattern, chart format and many other things.

Theoretical Elaborations of Technology Acceptance Model- TAM2 and TAM3

After testing the proposed model, a valuable next step is to attempt to gain a deeper understanding of the causal mechanisms that link system design features to user motivation. One way this can be accomplished is to begin to identify new variables that causally intervene between existing variables that have been observed to be related. A more detailed specification of the model is likely to generate further insights regarding user motivation, as well as provide the basis for more powerful diagnostic tools for evaluating proposed system designs. In this spirit, two model elaborations, referred to as TAM2 and TAM3 will be addressed.

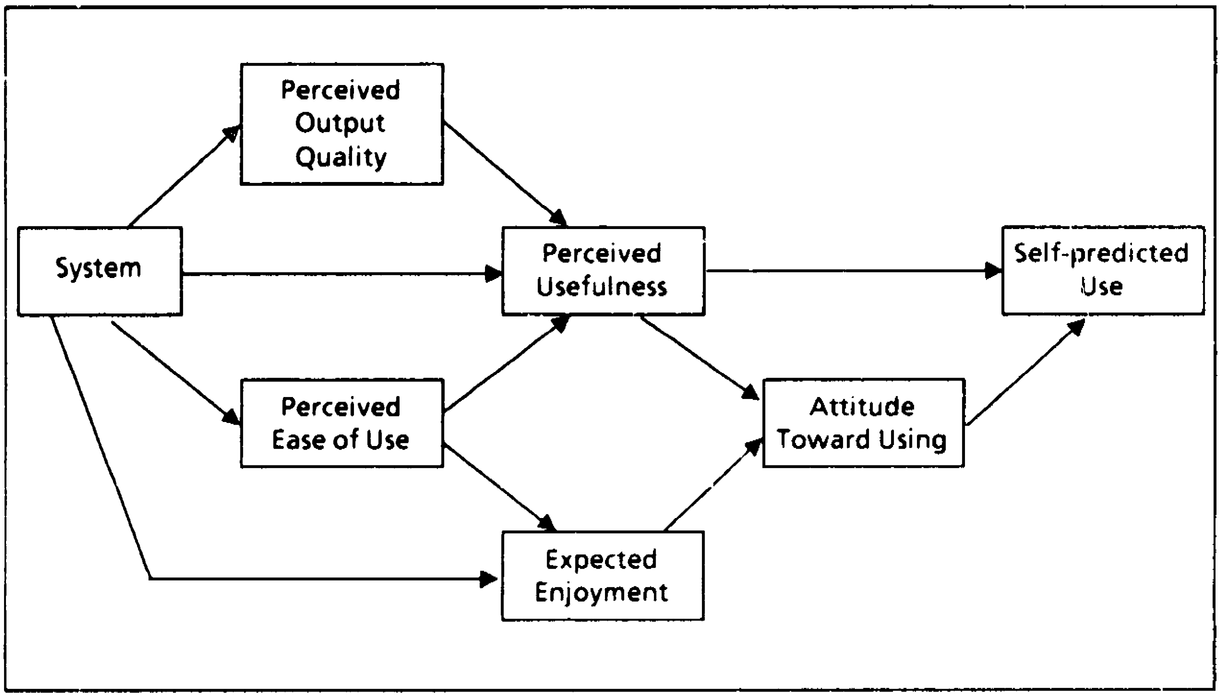
TAM2

Two additional variables are introduced and incorporated into the model to form TAM2: perceived quality of the output (QUAL), and anticipated enjoyment of using the system (FUN). The rationale for including these variables in the model and a specification of their relationship to the existing model variables will now be covered.

In the tests of the proposed TAM, using both the survey data (Davis, 1985) and the experimental data (reported in results section below), we observed that perceived ease of use (EOU) had a relatively minor effect on attitude (ATT) when perceived usefulness (USEF) is controlled for. Much of EOU's influence on ATT occurs indirectly through USEF, as found both in the survey (Davis, 1985) and the present experiment, as will be discussed later. Recall also that USEF is defined as "the extent to which a user perceives that using the target system in his or her job will increase his or her job performance". Together, these observations suggest that USEF may function as a "net" variable, reflecting considerations of both the perceived costs and benefits of using the system. EOU is defined as "the extent to which a user perceives that using the target system will be free of physical and mental effort." Thus, (when multiplied by -1) EOU may be regarded as a major part of the "cost" of using the system from the user's perspective. This cost-benefit interpretation raises two questions. First, if EOU relates to the "cost" of using the system, then how can we assess the "benefit" of using the system. Second, if USEF is truly a net variable, then why does EOU have a significant (albeit small) direct effect on ATT over and above USEF? These questions will be addressed in turn.

First, the perceived "benefit" of using a system may be difficult to define in general. Conceptually, we may think of the "benefit" of using the system as being related to the "product" of using the system, as distinct from the "process" of using it (the latter of which is expected to be strongly related to EOU). In trying to define and operationalize the benefit of using the system, we face two options: (1) to attempt to specify a general technique for measuring benefit which is applicable across a wide range of end-user tools, or (2) to identify and measure benefit idiosyncratically for given system domains. One of the desired characteristics of TAM is that it be a general model, capable of being applied across a wide range of end-user systems contexts. Unfortunately, the benefit, or product, of using a system would appear to vary

Figure 2. TAM2 Hypothesized Relationships



widely across system categories. For example, in the case of a database system, the product may be specific information elements, the benefit relating to certain properties it may possess (e.g. timeliness, accuracy), or may include "improved understanding of the business". Word processing systems may have various kinds of documents as their products. Decision support systems, modeling languages and spreadsheets may have "improved decisions" as their product. The product of electronic mail may be "faster communication". This diversity of "product" definitions raises significant problems for the general approach. Since in a general approach the benefit measure will need to be defined and worded in very broad and general terms, it is likely to be vague and ambiguous, which is likely to increase the measurement error. Moreover, the general approach may make it more difficult to assess benefit apart from EOU.

Hence, it will tend to become more of an overall evaluative construct, and may not exhibit discriminant validity from the existing ATT and USEF variables. This would reduce the diagnostic value of the benefit construct and obscure model testing efforts.

Therefore it appears that the more desireable approach is to model benefits in a context-specific manner. The present experiment deals with 2 business graphics systems intended for making visual aids for presentations. Thus, one may view the finished graphs themselves as the major product of using the systems. Note that this need not be true for graphics-based systems in general. For example, some systems use graphical interfaces as a front end to various functions for which the graphics become the means (i.e. process) and not the product per se. In the present experiment we are viewing the product to be the graphs produced, and hence we need to choose a way of evaluating the graphs. A major way of evaluating graphs is from the standpoint of their quality. Therefore, perceived quality of the output graphs (QUAL) is identified as a measure of the benefit of using the system. QUAL is theorized to intervene causally between System and USEF, as shown in Figure 2. That is, the quality of a chart is determined by the features of the system, and the quality of the charts in turn influences how useful the system is perceived to be. Note that a direct relationship from System to USEF is retained in TAM2. This was done because there is no basis for assuming that EOU and QUAL represent an exhaustive account of the cost-benefit considerations that are salient for evaluating these systems.

The second question is concerned with how the existence of a direct EOU-ATT effect can be reconciled with our view of USEF as a net cost-benefit construct. This direct link is an effect of EOU on ATT that is unrelated to USEF. USEF is a variable that places heavy emphasis on extrinsic motivation (e.g. "I am motivated to use the system in order to increase my performance for which I will be rewarded"). Recent theorizing has drawn attention to the fact that much human behavior is influenced by intrinsic as well as extrinsic motivation (Deci, 1975; Malone, 1981). Malone has applied this theory base to the design of enjoyable computer games. Deci (1975, p.23) gives this definition:

> "Intrinsically motivated activities are ones for which there is no apparent reward except for the activity itself. People seem to engage in the activities for their own sake and not because they lead to an extrinsic reward. The activities are ends in themselves rather than means to an end....the person is deriving enjoyment from the activity."

It is certainly plausible that computer use is at least partially intrinsically motivating. If so, this may account for the direct EOU-ATT link. This is the rationale for includingthe expected enjoyment of using (FUN) variable. EOU should influence FUN because a system which is easier to use may be more enjoyable to use, for example by increasing feelings of efficacy and competence (Deci, p. 55). Increased enjoyment of using should in turn increase attitude toward using (see Figure 2). In addition, we should expect a direct effect from System to FUN, since systems may have characteristics that make them more fun to use but which do not necessarily increase ease of use, by arousing challange, fantasy or curiosity (Malone, 1981), for example. The System-FUN-ATT chain may be the mechanism that produced the direct effect of system on attitude in the test of the basic TAM formulation as observed in the Survey, but not in the Experiment.

The TAM2 formulation is a recursive structural equation model that may be expressed in the following equations:

$$\begin{aligned}
EOU &= \beta_1 \text{ System} + e \\
QUAL &= \beta_1 \text{ System} + e \\
FUN &= \beta_1 \text{ System} + \beta_2 \text{ EOU} + e \\
USEF &= \beta_1 \text{ System} + \beta_2 \text{ EOU} + \beta_3 \text{ QUAL} + e \\
ATT &= \beta_1 \text{ FUN} + \beta_2 \text{ USEF} + e \\
BE &= \beta_1 \text{ USEF} + \beta_2 \text{ ATT} + e
\end{aligned}$$

Table 1 gives the hypothesized coefficients for this model. The above-

Table 1. TAM2 Constrained and Unconstrained Parameters

Dependent Variable	Independent Variable						
	S	E	Q	F	U	A	B
System	---	0	0	0	0	0	0
Perceived Ease of Use	β	---	0	0	0	0	0
Output Quality	β	0	---	0	0	0	0
Fun (Enjoyment)	β	β	0	---	0	0	0
Perceived Usefulness	β	β	β	0	---	0	0
Attitude toward Using	0	0	0	β	β	---	0
Behavioral Expectation	0	0	0	0	β	β	---

diagonal coefficients are assumed to be zero by definition of a recursive model. Statistical tests of whether the below-diagonal coefficients assumed to be zero are consistent with the experimental data will be performed.

TAM3

TAM3 makes a further elaboration on the structure of TAM2 by representing EOU and QUAL as task-specific, as opposed to general perceptions. The logic here is that how easy or hard a system is to use or the quality of its output (or benefit in general) may vary according to the nature of the task one is attempting to use the system to perform. Some end-user systems tend to be broad in their functional capabilities, encompassing many different kinds of tasks. They may be well suited to certain kinds of tasks, be able to do a passable job at others, and be totally unable to do yet other tasks. For example, Lotus 1-2-3 is primarily a spreadsheet tool. Accordingly, it performs spreadsheet functions effectively and easily. In addition, it can perform limited graphics and database functions, although its quality and ease of use would likely fall short of that of tools specifically intended for graphics and database tasks, respectively. In contrast, some systems address a rather limited functional scope effectively while totally omitting others (e.g. Automatic Teller Machines). Thus, when evaluating how easy to use a system is or how effective it's output is, greater diagnostic insight may accrue if we approach it on a task-specific level.

Another set of variables to be included in TAM3 are the importance or relevance (IMPORT) of specific tasks to a person's job. As an example, a system that is fairly easy to use and produces decent quality output for a narrow but important task domain may be perceived as more useful than both: (1) a system that is slightly more difficult to use, produces slightly lower quality output but can perform additional (but not very important) tasks, or (2) a system that produces higher quality output, is easier to use, but addresses tasks that are not very important. Therefore, we would expect the perceived usefulness of a system to be jointly determined by a system's perceived costs and benefits relative to a specific task domain, and the importance of that task domain to the individual. Since the importance of specific tasks are expected to vary from individual to individual, a task-specific aproach provides a framework for modeling USEF as being determined by the "fit" between the system and its capabilities on the one hand and the individual and his or her job requirements on the other. A way of expressing this mathematically is:

$$USEF_{system} = \Sigma_{task} \; (BENEFIT_{task,\, system} - COST_{task,\, system}) * IMPORT_{task}$$

The focus on identifying the importance of the task to the individual has the added advantage that it may permit us to identify groups or segments within the overall user population (e.g. writers, illustrators, analysts, planners) whose needs are fairly homogeneous, and for whom specific design configurations targeting those needs may be aimed.

The two systems being assessed in the experiment span two broad task categories: numeric graphs (e.g. line charts, bar charts, pie charts, etc.) and non-numeric graphs (e.g. diagrams, drawings, flowcharts, organizational charts). Thus, for the present experiment, TAM3 will address task-specific benefit (QUAL) and cost (EOU) for numeric and non-numeric graphs. As shown in Table 2 Chartmaster is specifically geared for numeric charts, and hence is expected to

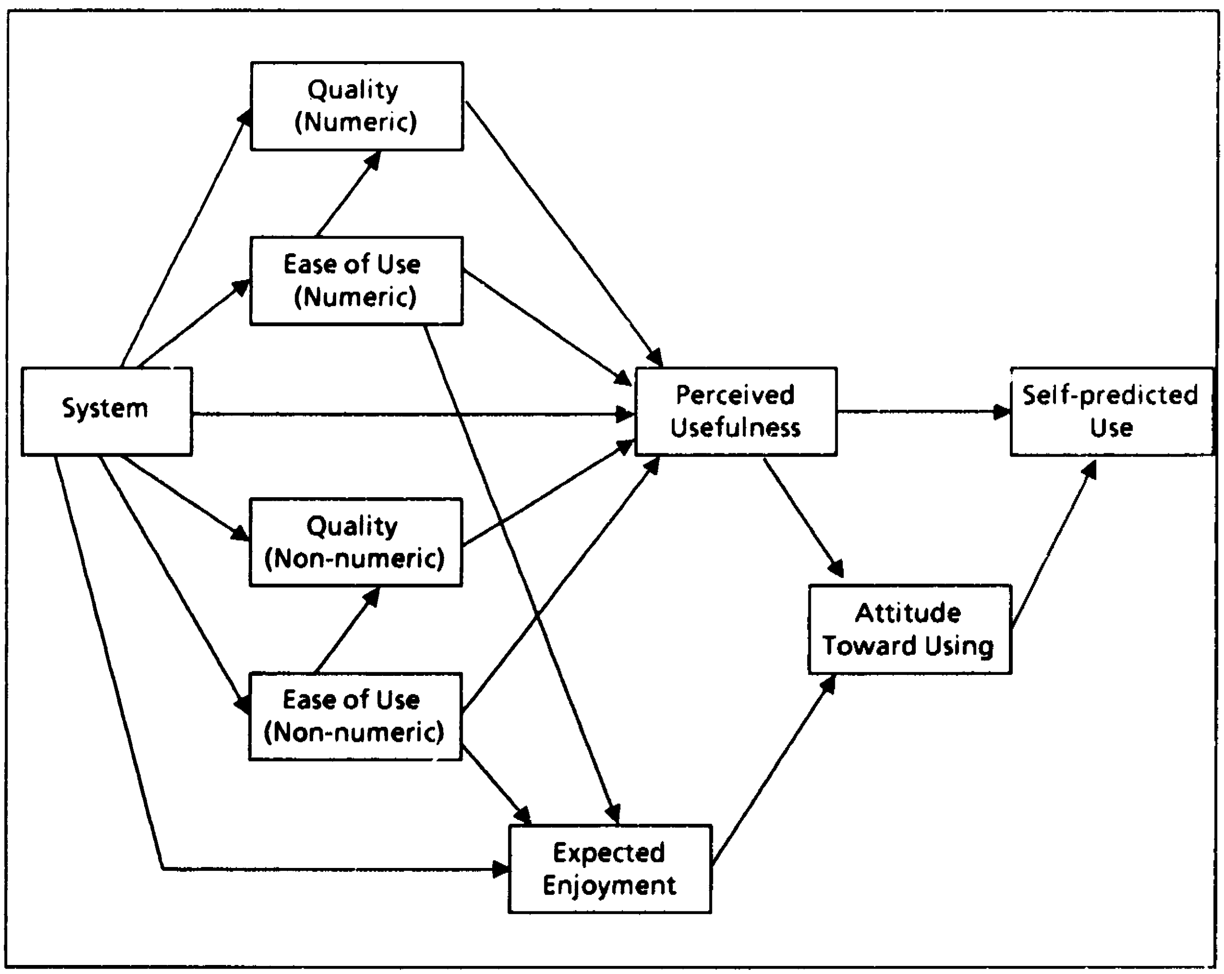
Table 2. Expected EOU and QUAL of System by Task Domain

System	Task Domain (Chart Type)			
	Numeric		Non-numeric	
	EOU	QUAL	EOU	QUAL
Chartmaster	+ +	+ +	0	0
Pendraw	+	+	+	+

be highly easy to use and produce good quality charts. Since Chartmaster is not technically able to make non-numeric charts, it's ease of use and quality are conceptually non-existent for this task domain. Pendraw, in contrast, is a broader product that enables the user to make both numeric and non-numeric graphs, but does not provide any specialized support for any specific type of graph. Therefore, its quality and ease of use is expected to be lower than Chartmaster for numeric charts and higher than Chartmaster for non-numeric charts.

Figure 3 specifies the hypothesized relationships among the TAM3 variables. The task-specific EOU and QUAL variables occupy the same causal position as

Figure 3. TAM3 Hypothesized Relationships



their general counterparts did in TAM2, with the same rationale. There is one exception, however. Notice that there is an effect of ease of use on quality. This is a result of a revision made to TAM2 based on the statistical analysis reported in the results section. In TAM2 there was an unexpected direct link from EOU to QUAL. In retrospect, this appears to be a result of the fact that people who find it more difficult to use Pendraw generate lower quality graphs. The same pattern would be expected on a task-specific basis, leading to the inclusion of this link. The multiplicative role of the task importance weights reflected in the above equation suggests that these importance weights should moderate the effect of EOU and QUAL on USEF for each task. Operationally, this implies that EOU and QUAL should each exhibit a significant interaction effect with IMPORT in determining USEF. Table 3 gives the pattern of hypothesized coefficients corresponding to TAM3.

Table 3. TAM3 Constrained and Unconstrained Parameters

Dependent Variable	Independent Variable								
	S	E1	E2	Q1	Q2	F	U	A	B
System	---	0	0	0	0	0	0	0	0
Ease of Use (Numeric)	β	---	0	0	0	0	0	0	0
Ease of Use (Non-num.)	β	0	---	0	0	0	0	0	0
Quality (Numeric)	β	β	0	---	0	0	0	0	0
Quality (Non-numeric)	β	0	β	0	---	0	0	0	0
Fun (Enjoyment)	β	β	β	0	0	---	0	0	0
Perceived Usefulness	β	β	β	β	β	0	---	0	0
Attitude toward Using	0	0	0	0	0	β	β	---	0
Behavioral Expectation	0	0	0	0	0	0	β	β	---

Method

Experimental Design

Participants were 40 Masters of Business Administration students from a large East Coast University who were given video and hands-on demonstrations of the 2 test systems according to the following counterbalanced within-subjects design:

Group 1:	PD_v O CM_v O	PD_h O CM_h O	n = 10
Group 2:	CM_v O PD_v O	CM_h O PD_h O	n = 10
Group 3:		PD_h O CM_h O	n = 10
Group 4:		CM_h O PD_h O	n = 10

PD_h = Pendraw Hands-on Demonstration

PDv = Pendraw Videotape Demonstration

CM_h = Chartmaster Hands-on Demonstration

CM_v = Chartmaster Video Demonstration

O = Measurement Observation (questionnaire)

Participants were assigned to one of the four treatment groups. Group 3 and 4 subjects received a hands-on demonstration only, and completed a separate questionnaire after interacting with each system. Group 3 subjects first interacted with Pendraw, and then Chartmaster. Group 4 subjects were exposed to the systems in the reverse order. Subjects in groups 1 and 2 were shown a videotape of each system, and completed a questionnaire after each videotape. Several days later they were given a hands-on demonstration of the two systems in the same order as the videotapes were shown (Pendraw first for Group 1 and Chartmaster first for Group 2).

This design has two characteristics that deserve mention. First, it is a counterbalanced design. Since the systems are shown in opposing order for different treatment groups, it is possible to separate effects due to order of presentation from effects due to the systems or the media by which the systems are demonstrated (videotape vs. hands-on). Prior research has observed that the order in which systems are shown may have a biasing effect favoring the first system shown (Ghani, 1981). Based on this prior research, we anticipate a significant order effect in the present experiment, which can be tested by virtue of counterbalancing. Further, since the order effect is an external stimulus, we hypothesize that the order effect will influence attitude and behavioral expectation indirectly via perceptions.

Second, the design has a control group (Groups 3 & 4) that does not receive the videotape treatment. There is some chance that having seen the videotape will carry over to the hands-on ratings for Group 1 and 2 subjects, thus contaminating them. This raises doubts about the validity of comparing the videotape and hands-on ratings for the same subjects. In order to rule out the existence of these carry-over effects, the hands-on ratings of Groups 1 and 2 may be compared to those of Groups 3 and 4, whose ratings have not had the opportunity to be contaminated by previous videotape viewing. A priori, we do not anticipate a significant carry-over effect because direct experience is known to be much more potent influence on attitude and perception formation than indirect experience (Fazio & Zanna, 1981).

Experimental Procedure

Subjects were recruited by making announcements in 4 courses at the participating university and circulating an announcement information sheet explaining the purpose and content of the study, with an attached sign-up sheet (see Appendix 2). Subjects signed up for the experiment by submitting the sign-up sheet along with a specification of time periods for which it would be convenient for them to participate. The sign-up sheets were used to prepare a schedule which was circulated among the participants. The videotape sessions were held in a classroom at the participating University. The hands-on sessions were held in the personal computer laboratory at the Sloan School of Management, MIT.

In the first session, (videotape session for Group 1 & 2 and hands-on for Group 3 & 4) subjects signed a consent form (See Appendix 3). The consent form explained the purpose and process of the experiment, informed subjects that they were free to withdraw at any time, contained a set of disclaimer points regarding the videotape, and asked subjects to agree not to discuss the experiment with other subjects until the study was over. In the videotape sessions, subjects were shown the two videotapes prepared by IBM for this experiment and asked to complete a questionnaire after each one.

In the hands-on sessions, subjects were first shown a booklet of sample charts created with the system they were about to be shown so they could get an understanding of the types of charts that could be prepared with the system. The sample booklet contained ten sample charts for each system. Seven of the sample charts were taken directly from the pages of the respective user manuals (and labeled as such). Three charts were created and printed using the Quadram Quadjet color inkjet printer (and labeled as such), so that subjects could understand the comparable output quality using the identical output device.

After viewing the sample booklets, subjects were given an instruction booklet for the system that they would be getting a hands-on demonstration of (Appendix 4). Subjects were asked to follow the instruction booklet step-by-step in order to become familiarized with the systems. The instruction booklets led the user through the different features of the systems and walked them through the creation of sample charts. The instruction booklets had been pre-tested on 4 non-sample subjects each to refine them and to calibrate the time required to complete them. Subjects were allowed approximately 45 minutes to complete the instruction booklets. The experimenter was present during the exercises to assist in case subjects ran into problems or questions; subjects were told to ask the experimenter for help if they ran into problems or questions. After using the first system, subjects were asked to complete a questionnaire (see Sample Questionnaire, Appendix 5). Then the process was repeated for the second system. After the second system, subjects were paid a $25.00 fee and thanked for their participation in the study.

Analysis of Potential Threats to Validity

In any research study there is a risk of drawing invalid conslusions from the observations obtained. As an effort to reduce this risk, the analysis reported below reviews a wide variety of known potential threats to validity from the standpoint of their possible role in ther present experiment. The intended benefits of this analysis are twofold. First, the recognition of likely threats to validity may lead to incorporating in the experiment precautionary procedures aimed at reducing or eliminating their effect. Second, in cases where such precautionary steps are impractical, the analysis will call attention to the possible existence of threats to validity so that interpretation of the experimental results may be made with such threats in mind.

The framework used for carrying out the validity analysis is that proposed by Cook and Campbell (1979). They define four categories of validity: internal, external, statistical conclusion, and construct. Cook and Campbell (1979, p. 50) define internal validity as the problem of deciding whether there is a causal relationship between two variables which have been observed to covary. In particular, attention is directed to the issue of whether some third variable may play a causal role in the relationship between the two variables of interest. The existence of such additional influences lead to rival explanations of the observed relationships between original variables, and if ignored, may lead to false positive or false negative conclusions. Cook and Campbell present a list of potential methodological artifacts that could play such a confounding role. External validity refers to "the approximate validity with which we can infer that the presumed causal relationship can be generalized to and across alternate measures of the same cause and effect and across different types of persons, settings, and times (Cook & Campbell, 1979, p. 37)." Statistical conslusion validity is concerned with possible false conclusions about whether or not two variables empirically covary (Cook & Campbell, 1979, p. 37). Note that whereas statistical conslusion validity is concerned with correctly detecting the existence of empirical covariation, internal validity is concerned with the causal interpretation of such covariation. Construct validity is concerned with the correspondence between experimental operations (manipulations and measurements) and the conceptual definitions of the underlying constructs that such operations are intended to represent (Cook & Campbell, 1979, p. 38).

Cook and Campbell (1979) explain that there are usually tradeoffs among the different types of validity, pointing out that "Some ways of increasing one kind of validity will probably decrease another kind." (p. 82), and that "...since some tradeoffs are inevitable, we think it unrealistic to expect that a single piece of research will effectively answer all of the validity questions surrounding even the simplest causal relationship." (p. 83). In view of this reality, it is important for researchers to attach priorities to the various kinds of validity. Such prioritization may be driven by the research objectives of the study. The present experiment is heavily oriented toward theory testing. Accordingly, we will follow Cook and Campbell's suggestion that "For investigators with theoretical interests our estimate is that the types of validity, in order of importance, are probably internal, construct, statistical conclusion, and external validity." (p. 83). Thus, we will carry out the validity analysis in this order. For each threat to validity identified by Cook and Campbell under each type of validity, the discussion will cover: how the threat is defined (employing frequent quotations of Cook and Campbell); its possible role in the present experiment; precautions aimed at reducing or ruling out the threat; and issues surrounding the impact of the potential threat on interpretations of results if the threat cannot be ruled out.

Internal Validity

History. This is a threat when "an observed effect might be due to an event which takes place between the pretest and posttest, when this effect is not the treatment of research interest." (Cook & Campbell, 1979, p. 51). In the proposed experiment, history would pertain to events that occur between the video and hands-on treatments for subjects in group 1 and 2. Of particular concern is the possibility that the student subjects may discuss among themselves the systems presented by video prior to the hands-on demonstration and evalutaion. Such discussion may influence subjects' perceptions and attitudes over and above the systems stimuli per se. Since the treatment of interest is the system, and not the social influence processes involved with opinion leadership etc., such social interaction represents a potential threat to validity. As a way of attempting to minimize the impact of this threat, subjects were asked to sign a consent stating that they agree not to discuss the contents of the experiment with any of their classmates until all of them had completed the entire experimental procedure.

Maturation. This is considered a threat when "an observed effect might be due to the respondent's growing older, wiser, stronger, more experienced, and the like between pretest and posttest and when this maturation is not the treatment of research interest." (Cook & Campbell, 1979, p. 52). In the proposed experiment, maturation may affect the hands-on based responses of subjects in Groups 1 and 2 as a result of their having previously seen the videotape. To the extent that such a carry-over effect is present, the hands-on responses of group 1 and 2 subjects should be regarded as being based on both video and hands-on presentations, rather than hands-on stimuli alone. In order to assess the extent of this threat, control groups 3 and 4, receiving hands-on demonstration only, were included in the experiment. If the hands-on based reactions of Groups 1 and 2 deviate substantially from those of Groups 3 and 4, the presence of a maturation (carry-over) effect will be considered plausible.

Testing. "This is a threat when an effect might be due to the number of times particular responses are measured." (Cook & Campbell, 1979, p. 52). This is especially pertinent where the response being measured is a performance variable, and repeated measurement may provide practice that would improve performance. In the present context, the variables being measured are perceptions and attitudes rather than performance. Thus, practice effects would fall into the category of Instrumentation validity threats, dealt with below.

Instrumentation. "This is a threat when an effect might be due to a change in the measuring instrument between pretest and posttest and not to the treatment's differential impact at each time interval." (Cook & Campbell, 1979, p. 52). In the present context, the same questionnaires are used for all treatment conditions. Nevertheless, the fact that subjects answer the same questions for up to four treatment combinations may lead to an experience effect that could effect the characteristics of the instruments. The error of the subject responses may be reduced with experience, due to greater familiarity with the scale. This is consistent with results from the survey which indicate higher reliability for the second administration of the questionnaire. Alternatively, error may be increased due to reduced concentration, boredom or fatigue. The fact that the administration of questionnaires will be separated in time should mitigate fatigue factors in the present design. Silk and Urban (1978) employed a similar repeated measures methodology in their "ASSESSOR" design, and ran a separate experiment which ruled out the presence of strong reactive effects due to repeated measurement. Moreover, the substantial attention given to the development of highly valid, reliable scales should make them generally robust, and hence less likely to be affected by repeated administration.

An additional issue related to instrumentation is that the responses measured may be sensitive to the order in which the treatment systems are presented. The system being shown first may provide a reference point or adaptation level against which the second system may be rated. As argued by Ghani and Lusk (1982), the first system shown may lead to the development of an "internal representation" (Newell & Simon, 1972) that is consistent with the first system. The second system, then, may be somewhat inconsistent with this representation. The lack of fit of the second system with the existing representation and/or the effort of re-organizing the internal representation to correspond to the new system may have the effect of negatively biasing one's atitudes and perceptions toward the second system. Thus the order of presentation and evaluation of the systems is a potentially significant instrumentation effect that should be dealt with. In the proposed design, order of presentation is counterbalanced so as to permit separating the bias due to order of presentation from the effect due to the research factors of interest: the difference between systems and the difference between the two media by which they are demonstrated.

Statistical Regression. "This is a threat when an effect might be due to respondents' being classified into experimental groups at, say, the pretest on the basis of pretest scores or correlates of pretest scores. " (Cook & Campbell, 1979, p. 52). In such a case, natural regression toward the mean of extreme scores may be confounded with treatment effects. This is not a concern in the proposed experiment because assignment of subjects to treatment groups was not based on pretest scores or correlates thereof.

Selection. "This is a threat when an effect may be due to the difference between the kinds of people in one experimental group as opposed to another. Selection is therefore pervasive in quasi-experimental research, which is defined in terms of different groups receiving different treatments as opposed to probabilistically equivalent groups receiving treatments as in the randomized experiment (Cook & Campbell, 1979, p. 53)." This threat is substantially mitigated in the present experiment via the probabilistic equalization of treatment groups due to random assignment of subjects.

Mortality. "This is a threat when an effect may be due to the different kinds of persons who dropped out of a particular treatment group during the course of an experiment (Cook & Campbell, 1979, p. 53)". This is a potential threat in the proposed experiment since subjects who have less favorable attitudes and perceptions toward the target systems may be more likely to drop out of the experiment between the video and hands-on treatments. If so, this could bias the comparison between the subjects receiving video (Groups 1 & 2) and those receiving hands-on only (Groups 3 & 4). In order to reduce attrition, as well as to increase subject motivation, subjects were paid $25.00 for their participation in the study. As will be seen in the results section, there was no attrition at all in the experiment.

Interactions with selection. Cook and Campbell point out that many of the above mentioned threats to validity may interact with selection. Since selection threats are substantially mitigated, so are their interactions with other threats.

Ambiguity about the direction of causal influence. "This particular threat is not salient in most experiments since the order of the temporal precedence is clear. Nor is it a problem in those correlational studies where one direction of causal influence is relatively implausible (Cook & Campbell, 1979, p. 54). " The present experiment and associated model validation has elements in common with both a purely experimental context and a purely correlational context. Since the system, medium, and order of presentation are being manipulated experimentally, it is clear that the direction of causal influence is from these factors to the response variables, and not vice versa. Aside from this, the specification of the causal relationships among the response variables is similar in spirit to a correlational study. Since none of them is being manipulated directly, the temporal precedence is not known. Theoretical arguments are used to assert the implausibility of particular causal orderings. The specification of causal relationships in the proposed models are based on a combination of well-established theories from social psychology (Bagozzi, 1982; Fishbein & Ajzen, 1975; Triandis, 1977), as well as specific theoretical arguments relative to the more narrow context within which the model is intended to be implied. Nevertheless, equivocality about the direction of causal influence may not be eliminated entirely, and we should adopt the stance of tentatively entertaining an approximate model and assessing the degree to which the observed data is consistent with it's proposed causal structure rather than attempt to use the data to "prove the correctness" of the specification. This is consistent with the philosophical position taken by the author regarding theory development and validation.

Diffusion or imitation of treatments. "When treatments involve informational programs and when the various experimental (and control) groups can communicate with each other, respondents in one treatment group may learn the information intended for others (Cook & Campbell, 1979, p. 54)." This would be a threat in the present experiment if the subjects receiving the video treatment (Groups 1 & 2) were to discuss the contents of the video with subjects who were to receive hands-on only (Groups 3 & 4), some of whom are classmates (as discussed under history, above). To reduce this threat as much as possible, subjects signed a consent form stating that they agreed not to discuss the contents of the experiment with their classmates until they had all completed the entire procedure.

Compensatory equalization of treatments. "When the experimental treatment provides goods or services generally believed to be desirable, there may emerge administrative and constituency reluctance to tolerate the focused inequality that results (Cook & Campbell, 1979, p. 54)." This threat is not applicable in the present context since there is little difference in the desireablity of the treatments and there is no sustained presence of the treatments in the subjects' everyday lives such that administrative and/or constituency intervention would be warranted.

Compensatory rivalry and resentful demoralization by respondents receiving less desireable treatments. "Where the assignment of persons or organizational units to experimental and control conditions is made public (as it frequently must be), conditions of social competition may be generated. The control group, as the natural underdog, may be motivated to reduce or reverse the expected difference (Cook & Campbell, 1979, p. 54)." "In an industrial setting the persons experiencing the less desireable treatment might retaliate by lowering productivity and company profits... (Cook & Campbell, 1979, p. 55)." These threats are not of particular concern in the present experiment since the treatments do not differ greatly in their desireability and response variables are not of a performance nature.

Construct Validity

Inadequate preoperational explication of constructs. "The choice of operations should depend on the result of a conceptual analysis of the essential features of a construct (Cook & Campbell, 1979, p. 64)." The emphasis here is on the careful specification of the meaning of proposed constructs prior to operationalization. "A precise explication of constructs is vital for high construct validity since it permits tailoring the manipulations and measures to whichever definitions emerge from the explications (p. 65)." A variety of response variables are used in the proposed experiment. Perceived usefulness and perceived ease of use were conceptually defined in Chapter 2, with operational measures derived from the definitions. Attitude toward using is a special case of the standard attitude toward the act variable from social psychology, with the definition and operationalization of the construct also taken from this reference paradigm (see Fishbein & Ajzen, 1975). The conceptual and operational definitions of the behavioral expectation construct are based on the work of Warshaw and Davis (1985; in press). In addition, three experimental factors are to be manipulated: system characteristics, order of presentation, and medium of presentation. Due to their many differences, comparing Pendraw vs. Chartmaster is clearly an example of manipulating "system characteristics". Order of presentation is a methodological control factor that has no surplus conceptual connotation. Medium of presentation refers to the comparison of videotape presentation with hands-on interaction. Explication of these categories suggested that the videotape and hands-on treatments should be operationalized so as to be as representative as possible of how these media would each be used in practice. The particular video and hands-on stimuli used are examples of their respective categories. Thus, for all three research factors, a fairly direct translation from construct explication to operationalization was possible.

Mono-operation bias. "Many experiments are designed to have only one exemplar of a particular possible cause, and some have just one measure to represent each of the possible effect constructs. Since single operations both underrepresent constructs and contain irrelevancies, construct validity will be lower in single exemplar research than in research where each construct is multiply operationalized in order to triangulate on the referent (Cook and Campbell, 1979, p. 65)." Validated multi-item measures are used to measure most of the response constructs, for, as Cook and Campbell (1979, p. 65) point out: "There is rarely an adequate excuse for single operations of effect constructs, since it is not costly to gather additional data from alternative measures of the targets." Unfortunately, limitations of the experimental setting prohibited the replication of the causal constructs. In the case of order of presentation, it is logically impossible to manipulate this factor in any other way. Two systems were compared against one another. To increase this number would either increase the sample requirements substantially or increase the time required to expose subjects to all systems beyond a reasonable time period. The parallel situation exists in the case of the hands-on vs. videotape manipulation: the production of more than two videotapes was beyond the resource limitations of the experimental setting. These are typical constraints facing researchers, as Cook and Campbell (1979, p. 65) explain: "There is more excuse for having only one manipulation of a possible causal construct. This is because increasing the total number of treatments in a factorial design can lead either to very large sample research or to small sizes within each cell of the design should it not be possible to increase the total sample size."

Mono-method bias. "To have more than one operational representation of a construct does not necessarily imply that all irrelevancies have been made heterogeneous. Indeed, when all the manipulations are presented in the same way, or all measures use the same means of recording responses, then the method is itself an irrelevancy whose influence cannot be dissociated from the influence of the target construct (Cook & Campbell, 1979, p. 66)." First let us consider the causal constructs of the proposed experiment. The order of presentation of systems was reversed for half of the subjects, thus providing alternative methods. As discussed above, the order of presentation is expected to have a systematic effect on subject responses. In a sense, the two different media of presentation employed, videotape and hands-on, represent alternative methods of presentation. However, this manipulation is a experimental factor of interest, and was not done for the purposes of reducing methodological irrelevancies. As discussed under the heading of mono-operation bias, replication of the method of presenting stimuli would have been costly in terms of sample size. Although not planned, an additional method of presentation within the hands-on treatment, for example would have allowed us to rule out the possiblity that the observed response effects may be due to some methodological idiosyncracy of the particular hands-on treatment used.

Regarding the measurement of response variables, Cook and Campbell (1979, p. 66) recommend that researchers give thought to: "(a) using methods of recording other than paper-and-pencil, (b) varying whether the attitude statements are positively or negatively worded, or (c) varying whether the positive or negative end of the response scale appears on the right or left of the page." Methods other than paper-and-pencil are somewhat underdeveloped for the class of variables being addressed in the present research: attitudes and perceptions. Observational techniques are difficult to apply since the constructs of interest are internal to the subjects' minds and relatively inaccessable to outside observers (although such non-verbal cues as body language and facial expressions can be indicative). Researchers have begun to address the potential for using physiological observations such as galvanic skin response as indicants of attitudes, although the use of such methods would be premature a this point in their development. Projective techniques, such as TAT and sentence completion also show promise from the standpoint that they seem geared to tap the unconscious aspects of subject's perceptual and attitudinal structures. The process of tailoring such techniques to the task of measuring the specific variables of interest in the present context is beyond the intended scope of the research. Computer-based response recording is another possibility, but runs the risk of confounding the measurement process with one of the treatments. This may occur, for example, if the computer response system resembled one of the treatment systems more than the other, or if subjects begin to rate the computer response system itself rather than the target system. The paper and pencil response scales employed are easy to use by non-experts and therefore would appear more appropriate for the applied user acceptance testing context for which the model is being developed. Nevertheless, the research would no doubt be improved by using alternate recording media.

The measurement scales for many of the variables employed reversed scale polarities for the items. As will be seen, this does not seem to have hampered their reliability and validity. In the survey reported in Chapter 5, a mixture of positively and negatively worded items was used in measuring the ease of use construct. The fact that the negatively worded items exhibited much lower inter-item correlation than the positively worded ones suggests that negative wording increased the measurement error. Based on this observation and as well as the literature on direction of wording effects (Campbell, Siegman & Rees, 1967) it was concluded that the increase in measurement error does not appear to be compensated by an increase in validity, and hence the perceptual items used in the present experiment are all framed positively.

Hypothesis-guessing within experimental conditions. This refers to the possibility that subjects may attempt to guess what the hypotheses of the experiment are and attempt to modify their behavior accordingly, either to confirm or disconfirm the hypothesis. In the present research, this issue is of concern in two areas. Subjects may attempt to guess which system is hypothesized to be more highly rated or they may assume that there is a connection between their video and hands-on responses. Neither systems was hypothesized to be more favorable rated a priori. Thus, the treatment stimuli were constructed so as to be as objective as possible toward the systems. This tends to increase the possibility that whichever system a subject guesses to be the hypothesized favorite (assuming the subject is engaging in hypothesis-guessing) is the one they themselves favor, (although this need not always be the case, for example a subject may reason: "the hypothesis must be that system A is the favorite since most people would prefer A, even though I prefer B"). If hypothesis-guessing subjects answer in accordance with their guesses, therefore, this would lead to responses which are consistent with their own evaluations, and the threat to validity is reduced. The risk always exists, however, that hypothesis-guessing subjects may answer in such a way as to contradict their guessed hypothesis, which would represent a greater threat to validity. Fortunately, this is thought to be more likely under conditions of resentful demoralization, which, as discussed above, are not expected to be present in this experimental context. As a further step toward reducing hypothesis-guessing, subjects are asked to rate each system after it has been used, which reduces comparative analyses that may encourage hypothesis-guessing.

Subjects in Groups 1 and 2 may guess that a research hypothesis relates to the ability to predict hands-on responses from video responses. If so, they may attempt to make their hands-on ratings consistent with their video ratings. Two factors mitigate this potentiality. First, since the hands-on rating task is performed approximately a week after the video rating task, subjects should have difficulty remembering their video responses. Second, other research (e.g., Nisbett & Wilson, 1977) has shown that when subjects undergo changes in beliefs or attitudes, they are often unaware of the changes, and will tend to incorrectly infer that their previous ratings were the same as their present ratings. If so, then subjects may believe that their hands-on ratings are consistent with their video ratings in cases where they are not. The additional possibility exists that subjects may assume that the purposes of the experiment are other than as they appear, and attempt to guess hypotheses that do not really exist. Since efforts to act in accordance with these extranneous hypotheses may become confounded with the experimental effects, such hypothesis-guessing was discouraged by explaining in the consent form which all subjects signed that no deception is involved in the experiment. All of these concerns are based on the assumption that subjects engage in hypothesis-guessing. There is actually little evidence that such behavior is pervasive (Cook, 1967; Weber & Cook, 1972).

Evaluation apprehension. "Rosenberg (1969) has reviewed considerable evidence from laboratory experiments which indicates that respondents are apprehensive about being evaluated by persons who are experts in personality adjustment or the assessment of human skills. In such cases respondents attempt to present themselves to such persons as both competent and psychologically healthy (Cook & Campbell, 1979, p. 67)." Such effects are most likely when the response of interest is some evàluation of the subject, such as performance or intelligence. In the present experiment, the focus was on having the subjects evaluate the target systems, rather than having the experimenter evaluate the subjects. This should reduce the perceived need for, and measurement impact of, subjects' presenting themselves in a favorable light. The perceived ease of use variable, however, may be interpreted by subjects both in terms of the system's characteristics and their own abilities. Subjects may upwardly bias their ease of use ratings in order to reflect favorably on their own abilities. Two precautions were taken to minimize this effect. First, in the consent form, subjects were informed that "we are evaluating the graphics software itself, and not you, the participants." Second, although the experimenter was present and available for questions during the procedure, he was directing his attention primarily to his paperwork, and not to the behavior of subjects attempting to use the system.

Experimenter expectancies. "There is some literature (Rosenthal, 1972) which indicates that an experimenter's expectancies can bias the data obtained. When this happens, it will not be clear whether the causal treatment is the treatment-as-labeled or the expectations of the persons who deliver the treatments to the respondents (Cook & Campbell, 1979, p. 67)." Since the experimenter for this experiment is the present author, much of the discussion of hypothesis-guessing is pertinent here. Since the experimenter had no prior expectations regarding which system would be favored, there were no expectations to guide subjects reactions. Even if subjects were apprised of the experimenters expectancies toward the video vs. hands-on manipulation, subjects would find it difficult to comply with the expectations due to reasons discussed in the context of hypothesis-guessing above. There is little opportunity for experimenter expectancies to be transmitted socially insofar as the experimental treatments were largely free of social interaction with the experimenter (i.e., watching videotapes, following instruction booklets, and filling out a questionnaire).

Confounding constructs with levels of constructs. This a problem if a researcher erroneously concludes that a causal construct has no effect on a response variable when the true case is that such a causal relationship does exist but the experimental manipulation was not strong enough to reveal the relationship. This is an example of the problem of accepting the null hypothesis and is dealt with further under the section on statistical conclusion validity below. There is a risk in the present experiment that the two systems do not provide a strong enough manipulation to observe significant differences in response variables. The fact that the two systems differ along many characteristics offers hope that this manipulation will be strong enough.

Interaction of different treatments. "This threat occurs if respondents experience more than one treatment which is common in laboratory research but quite rare in field settings. We do not know in advance whether we could generalize any findings to the situation where respondents received only a single treatment. More importantly, we would not be able to unconfound the effects of the treatment from the effects of the context of several treatments (Cook & Campbell, 1979, p. 68)." This is a pertinent concern in the present research context since a repeated measures design is employed. In the case of the video vs. hands-on treatment, a hands-on only control group was employed. This will permit an analysis of the possible carry-over effects of the video treatment into the hands-on treatment. However, all subjects are exposed to both systems. Thus, the findings will be relevant to the context of the both systems, and the ability to generalize them to contexts in which more or different alternative systems are available is unknown. Since we would naturally expect that potential users would have a variety of alternative systems available for their use, we are less interested in generalizing to the "single treatment" case. Rather, our concern centers around the sensitivity of the responses to the particular alternative set tested. Another way of viewing this is to say that evaluation tasks may be sensitive to the salient "choice set". From a behavioral prediction point of view, we would expect the ability to predict a subject's behavior from their response ratings to be at a maximum when the choice set of the subject's behavioral response is in correspondence with the choice set within which their ratings were taken (in this sense, choice set may be seen as another element of attitude specificity along with target, action, context, and time, as defined by Ajzen & Fishbein (1980)). Since we are at present uncertain about the sensitivity of ratings to the choice set, we would be prudent in generalizing the experimental findings to the specific choice set employed, and not to other non-tested choice sets.

Interaction of testing and treatment. The concern here is whether pre-testing and/or repeated posttesting may become confounded with the treatment so that causal results may not be generalized to other testing contexts where a different pattern of tests may be used. In the present research, this would not appear to be a major concern. First, the tests (measurements) are not expected to be highly confounded with the treatments in the first place. Second, the pattern of tests is representative of the target research context to which we would like to generalize the results, namely, applied user acceptance testing.

Restricted generalizability across constructs. The point being made here is that by selecting one or a small set of response variables, one limits the ability to generalize the findings to additional response variables that may be of interest to other researchers, since "Sometimes treatments will affect dependent variables quite differently (Cook & Campbell, 1979, p. 68).". This is dealt with in the present experiment by measuring an array of response variables that are indeed expected to be affected differently by treatments and which are hypothesized to relate to one another in a specific theoretical model. In approaching the research in this way, it is hoped that a more complete picture of the treatment effects will be possible than by employing a single dependent varaible.

Statistical Conclusion Validity

Low statistical power. "The likelihood of making an incorrect no-difference conclusion (Type II error) increases when sample sizes are small, and α is set low (Cook and Campbell, 1979, p. 42)." For the present research context, α will be set at .05 for determining whether an effect is significant, which is a generally accepted criterion and not so low as to drastically inflate the probability of Type II error. Nevertheless, if an effect is nonsignificant, we would not be able to unequivocally "accept the null hypothesis". Strictly speaking, we should only conclude that we are unable to reject the null hypothesis, since it may be the case that a true effect actually exists but it's magnitude is so small that the statistically tests used were unable to detect it. This observation has several implications for the present research. First, we should avoid conclusive interpretations of non-significant results. Second, we should base interpretations on the estimated magnitude of the effects (point and interval estimates) rather than strictly relying on their statistical significance or lack thereof. Finally, in regression analyses performed for the purpose of model testing, independent variables hypothesized to be significant but found to be non-significant should be left in the regression equation. We will assume that their effects are simply not large enough to be detected, and draw interpretations based on the estimated magnitude of their effects. Removing them from the regression would open the possibility of biasing the coefficient estimates for the remaining independent variables.

Violated assumptions of statistical tests. "Most tests of the null hypothesis require that certain assumptions be met if the results of the data analysis are to be meaningfully interpreted. Thus, the particular assumptions of a chosen statistical test have to be known and - where possible - tested in the data on hand (Cook & Campbell, 1979, p 42)." Although it is difficult to anticipate the nature and impact of departures from such assumptions a priori, the experimental data will permit a check of many of the key assumptions.

Fishing and the error rate problem. "The likelihood of false concluding that covariation exists when it does not (Type I error) increases when multiple comparisons of mean differences are possible and there is no recognition that a certain proportion of the comparisons will be significantly different by chance (Cook & Campbell, 1979, p. 42)." As in the case of low statistical power, this threat can be mitigated by shifting emphasis to the estimated magnitude of the effect rather than making strict "yes or no" interpretations based on the significance of the test statistic. One context in which we are especially interested in avoiding false positives is in the test of the causal models. Here we must test that causal relationships hypothesized to be nonexistant are observed to be nonsignificant in the corresponding regression equations. We may use hierarchical regression and associated F-tests to jointly test the hypothesis that all of the specified coefficients are equal to zero. Further, across all of the regressions corresponding to a model, we may employ the χ^2 statistic given by Land (1973) to simultaneously test whether the set of causal relationships hypothesized to be non-significant is consistent with the data.

The reliability of measures. Since unrelibility inflates the standard errors of estimates, unreliable measures cannot be depended on to exhibit true effects. This is especially crucial in the present context since many variables must be measured, and their interrelationships tested in order to assess a theoretical model. For that reason, a great deal of emphasis was placed on the development of highly reliable valid measures (see Chapter 4) which substantially eliminates this threat to validity for the major analyses of the experiment. Some additional variables are introduced in the experiment for the purpose of exploring theoretical elaborations of the proposed model. Some of these employ multi-item scales, the reliability of which will be estimable based on the experimental data. Some of the new variables will be measured with single-item scales, due to the need to constrain the length of the questionnaire and in light of the exploratory nature of these variables. It will be kept in mind that interpretations based on these unreliable single-item measures are tenuous.

The reliability of treatment implementation. "The way a treatment is implemented may differ from one person to another if different persons are responsible for implementing the treatment. There may also be differences from occasion to occasion when the same person implements the treatment. This lack of standardization, both within and between persons will inflate error variance and decrease the chance of obtaining true differences (Cook and Campbell, 1979, p. 43)." The videotape treatments are well standardized due to their fixed nature. In an effort to make the hands-on interaction treatments as standard as possible from one subject to another, it was decided to use a self-administered instruction booklet rather than employing a human tutor. The instruction booklet is much more fixed across occasions than a human tutor would be.

Random irrelevancies in the experimental setting. "Some features of an experimental setting other than the treatment will undoubtedly affect scores on the dependent variable and will inflate error variance. This threat can be most obviously controlled by choosing settings free of extranneous sources of variation or by choosing experimental procedures which force respondents' attention on the treatment and lower the salience of environmental variables

(Cook & Campbell, 1979, p. 44)." To a large extent, both of these suggestions will be followed. The video and hands-on treatments will both be conducted in the same respective locations. The only major variable expected to vary is the number and identity of the specific other participants present during the procedure. The effect of this is minimized since subjects will have agreed not to discuss the experiment with other participants, and the procedure is such that it focusses subjects' attention on the treatments themselves.

Random heterogeneity of respondents. "The respondents in any of the treatment groups of an experiment can differ on factors that are correlated with the major dependent variables (Cook & Campbell, 1979, p. 44)." Assuming such variables do not differ across treatment groups (as in the case of random assignment) their effect is to inflate the error variance. Since the participants in this experiment are MBA students, it is anticipated that they will be quite hetergeneous, with some of the factors that differentiate them being related to the response variables of interest. Although this heterogeneity has the effect of inflating error variance, it greatly expands the generalizability of the results, which is a matter of external validity.

External Validity

Interaction of selection and treatment. This has to do with the categories of persons to whom the treatments may be generalized. As discussed above, it is expected that the MBA student subjects will be a fairly heterogeneous sample. However, the selection process involved recruiting volunteers, which suggests that we may have a group of people who are generally more positively disposed toward systems of the type dealt with in the experiment. Thus, the responses of the sample subjects may be more positive on average than those of the non-sample subjects from their parent population. However, this does not imply that subjects are expected to be differentially more favorable to one or the other of the test systems than non-sample subjects. Hence it is plausible that the magnitude and direction of the differences in response between systems would be similar for the two groups. Aside from the issue of what the ratings of the specific systems are, other causal relationships may be more plausibly generalized beyond the sample group. The relative importance of ease of use and usefulness in influencing attitude, for example, may be a comparatively more robust finding that is less sensitive to the idiosyncrasies of the specific sample group.

Interaction of setting and treatment. "Can a causal relationship be obtained in a factory be obtained in a bureaucracy, in a military camp, or on a university campus? The solution here is to vary settings and to analyze for a causal relationship within each (Cook & Campbell, 1979, p. 74)." As alluded to above, the fact that the MBA participants are expected to be from a variety of work backgrounds will enable the present exeperiment to be generalized moreso than if the subjects were all drawn from a narrow homogeneous environment. Still, external validity will accumulate as additional settings are studied. One important aspect of the setting that will need to be varied across studies is the nature of the specific systems addressed. Since it is quite difficult to employ a large number of systems in a single experiment, the most effective route to external validity appears to be to replicate the experiment in a variety of contexts using different systems. This course is already being followed in that the survey reported in Chapter 5 was conducted in a different setting and dealt with 2 different systems than the present experiment.

Interaction of history and treatment. "To which periods in the past and future can a particular causal relationship be generalized? (Cook & Campbell, 1979, p. 74)." A major historical phenomenon at issue in this research context has to do with the existing predominant technological environment of the subjects. As time moves forward and subjects become acquainted with ever more advanced systems, their frame of reference for evaluating specific target systems is likely to shift. This is a pertinent concern to the extent that in conducting user acceptance tests we are interested in predicting likely acceptance patterns of systems at some future time. Thus, to the extent possible, we should anticipate the future scenario and reflect its characteristics in the choice set to be examined in a given acceptance test. To some extent this has been accomplished in the present experiment. The target systems are likely to be more advanced than would be commonplace for the subjects at the present time, but are likely to be commonplace within a five-year time horizon.

Results

Summary

Validation of Videotape as a Substitute for Hands-on Interaction

(1) Presenting new systems to potential users by videotape appears to enable subjects to form accurate *attitudes*, *usefulness* perceptions, *quality* perceptions and *behavioral expectations* (self-predictions of use) regarding the new systems, as compared to hands-on interaction.

(2) Videotape presentation does not appear to enable subjects to form accurate perceptions regarding how *easy to use* a new system will be, as compared to hands-on interaction.

(3) The inaccuracy of *ease of use* perceptions based on videotape presentation was different for the two systems tested: Pendraw & Chartmaster. The videotape-based ease of use perceptions overstated the hands-on-based ease of use perceptions for Pendraw and understated them for Chartmaster.

(4) The order in which systems are presented to participants was a significant factor that influenced responses to all motivational variables analyzed: perceived ease of use, perceived usefulness, attitude toward using, and behavioral expectation, with the advantage going to whichever system was shown first. Moreover, this effect was much more pronounced for Pendraw than for Chartmaster. Presentation order was controlled for in the present study, and should be controlled for in future user acceptance testing applications.

(5) Subjects' average confidence in their ratings was not significantly different between ratings based on video exposure only and those based

on hands-on exposure only. However, subjects having previouslly seen the videotape reported significantly higher mean confidence in their hands-on ratings than either their own video-based ratings or the hands-on ratings on the non-video control group. This suggests that total exposure time, and not medium (video vs. hands-on), influences subjects confidence in their ratings.

(6) Subject's confidence in their ratings appears to have some value in determining the extent to which individuals' video-based ratings will predict their hands-on ratings, although further research would be needed to gain enough of an understanding of the confidence variable in order to apply it diagnostically in practice.

Further Validation of Technology Acceptance Model (TAM)

(1) The experimental data substantiated the theoretical causal structure of the Technology Acceptance Model (TAM). Especially pertinent is the dominant role for perceived usefulness, which has a powerful effect on attitude toward using and a powerful direct effect on self-predicted usage behavior above and beyond its indirect effect through attitude toward using. Also of interest is the further confirmation of the comparatively limited role of perceived ease of use, which has a smaller influence on attitude toward using than does usefulness, and no direct effect on behavior beyond this indirect effect via attitude. The causal analysis enabled us to estimate that usefulness was 2.65 times as important as ease of use in determining self-predicted system usage in the experiment. This compares to a 1.46 usefulness-ease of use importance ratio found in the previous user survey.

Evaluation of Specific Systems Tested

(1) Based on hands-on-based ratings, Chartmaster was rated as superior to Pendraw across all response variables except two. Chartmaster was rated as significantly better than Pendraw in terms of: overall ease of use, overall quality of output, usefulness, attitude toward using, and self-predicted use. On a task-specific level, Chartmaster was significantly favored for ease of use and quality in making *numeric* graphs. Pendraw rated more favorably for ease of use and quality for *non-numeric* graphs, although these differences were not statistically significant.

Theoretical Elaboration of TAM

(1) Two additional variables: perceived quality of the output and expected enjoyment of using the system were incorporated into the TAM formulation, leading to an alternative model referred to as TAM2. This alternative model was largely supported by the experimental data, with the exception of one causal relationship that was hypothesized to be insignificant but which turned out to be significant: the effect of ease of use on quality. Further analysis supported the interpretation that people who find Pendraw easier to use are able to create higher quality graphs than those who find it more difficult to use. The data revealed a more influential role for perceived quality than for anticipated enjoyment in influencing motivation to use a system.

(2) A third formulation, TAM3, broke the ease of use and quality perceptions down to a task-specific level: numeric vs. non-numeric charts. The data confirmed expectations that Chartmaster would be easier to use and would produce higher output for numeric graphs while Pendraw should be better in these areas for non-numeric graphs. The model also incorporated the perceived importance of each of these graph types for the individuals' jobs, finding that on the average numeric graphs are

perceived to be more important than non-numeric graphs. Although at the very earliest stages of develoment, TAM3 appears to provide a potentially powerful framework for: (1) carrying out a finer-grained diagnostic analysis of system functionality, (2) understanding the "fit" between a system and its capabilities on the one hand and users' job needs on the other, and (3) identifying segments of users whose needs are relatively homogeneous and for whom specific system configurations may be targeted.

Description of Sample

The sample of participants consisted of 40 Masters of Business Administration Students, 24 males and 16 females, from a large East Coast U.S. university. Table 4 shows how subjects rated their own computer experience

Table 4. Subjects Prior Computer Experience

Prior Computer Experience	Percent of Subjects			
	None	Limited	Moderate	Extensive
Computers in general	0.0	35.0	47.5	17.5
Personal Computers	17.5	17.5	50.0	15.0
IBM PC	39.5	15.8	26.3	18.4
CHART-MASTER	95.0	5.0	0.0	0.0
Other software similar to CHART-MASTER	60.0	30.0	7.5	2.5
PENDRAW	95.0	5.0	0.0	0.0
Other software similar to PENDRAW	90.0	10.0	0.0	0.0

across a number of categories. The majority of the participants (33) reported having "limited" to "moderate" experience with computers in general, while 7 reported having "extensive" computer experience. Regarding personal computer experience, 27 participants reported having "limited" to "moderate" experience, with 7 and 6 participants reporting "none" and "extensive", respectively.

Subjects reported having had an average of 4.79 years of full time work experience. Eleven of the subjects had between 0 and 2 years of full time work experience, 21 had between 3 and 7 years experience, and 8 reported having more than 7 years. The industries that subject's have worked in (or anticipate working in) span a wide range including: Financial (11); Health (7); Education (4); Government (4); Manufacturing (3); Retail (2); Services (1); and Transportation (1). Six subjects rated "other" as their industry affiliation and one person did not respond at all.

Psychometric Properties of Scales

Four major variables were measured for the purposes of validating the proposed model and testing the viability of substituting video: (1) perceived ease of use (EOU), (2) perceived usefulness (USEF), (3) attitude toward using (ATT), and (4) self-predicted usage, or behavioral expectation (BE). Table 5

Table 5. Cronbach Alpha Reliabilities of Scales Used

			By System & Medium			
			Pendraw		Chartmaster	
Variable	Number of Items	Alpha Overall	Video	Hands-on	Video	Hands-on
EOU	6	.93	.85	.92	.93	.82
USEF	6	.97	.96	.98	.93	.93
ATT	4	.95	.97	.97	.93	.86
BE	2	.93	.95	.92	.75	.91

shows the number of measurement items used to measure each of these four variables, and the reliability coefficients obtained. The measures used to address EOU, USEF, and ATT are refinements of those developed and tested by the author and reported in Davis (1985) (see sample questionnaire Appendix 5). The BE measures are based on recently published research (Warshaw & Davis, 1985). The overall Cronbach alpha reliabilities are all above 0.90, thus providing strong evidence of reliable measures. Among the disaggregated alpha coefficients, only one coefficient fell below .80 - the reliability of BE for Chartmaster under the video treatment (alpha of .75). These dissagregated alpha estimates are based on only 10 individuals, however, and are therefore subject to error. Each measurement question (item) was asked on a seven-point rating format, which was coded from 1 to 7 where 7 was the highest rating. The coded values for the individual items were averaged together to form the scale values for each variable.

Eight additional variables were measured for purposes of testing theoretical elaborations to the existing model: perceived quality of the output (QUAL; 3 items; Cronbach alpha = .70), expected enjoyment of using the system (FUN; 3 items; alpha = .92), task-specific ease of use and quality measures for numeric and numeric charts (EOU_{num}, EOU_{non}, $QUAL_{num}$, $QUAL_{non}$; 1 item each), and Importance measures for numeric and non-numeric charts ($IMPORT_{num}$ and $IMPORT_{non}$; 2 items each; alpha = .77 and .89 respectively). Finally, single-item confidence measures were taken for all of the above variables with the exception of the task-specific constructs. The confidence measure is based on the work of Fazio & Zanna (1981).

Validation of Videotape as a Substitute for Hands-on Interaction

We now will address whether video demonstration can serve as an accurate substitute for hands-on interaction. This is done by assessing the effect of medium (videotape presentation vs. hands-on interaction) on the response variables. Since we hypothesized that the system (Pendraw vs Chartmaster) and order of presentation will have a significant effect on the response variables, they are included in the analysis of the medium-response relationship in order to control for their potentially confounding effects. Table 6 gives the mean

Table 6. Mean Ratings by System by Variable by Treatment Group

System	Variable	Treatment Group: Videotape (Group 1& 2) Shown First	Videotape (Group 1& 2) Shown Second	Hands-on (Group 1 & 2) Shown First	Hands-on (Group 1 & 2) Shown Second	Hands-on (Group 3 & 4) Shown First	Hands-on (Group 3 & 4) Shown Second
Pendraw	EOU	6.25	5.60	5.35	4.28	5.52	4.02
	USEF	4.82	3.48	5.17	2.75	4.02	3.18
	ATT	5.43	4.68	5.25	3.82	4.88	4.03
	BE	4.65	3.45	5.30	2.90	3.95	2.75
Chart-	EOU	5.67	5.47	6.73	6.17	6.40	6.37
master	USEF	4.60	5.48	5.68	5.68	5.30	5.03
	ATT	5.98	5.80	5.98	5.85	6.05	5.35
	BE	4.85	5.55	5.70	5.50	5.55	5.50

Note: Ratings may range from 1 (low) to 7 (high).

values for the TAM variables broken down by system (Pendraw vs. Chartmaster), and treatment group (Videotape, Hands-on after having seen video tape, and hands-on without having seen videotape).

Test of Video Carry-over Effect

For maximum precision, we would like to employ the hands-on responses for all four treatment groups in the analysis. However, we must first rule out the possibility that having previously seen the videotapes may have contaminated the hands-on responses of Groups 1 & 2. This may occur, for example, if the perceptions and attitudes formed based on the videotapes were to "carry over" into the hands-on ratings. We can assess the extent of such contamination by testing the effect of having seen the video on subsequent hands-on responses. We hypothesize that having seen the video will not significantly affect subsequent hands-on responses. *System* (Pendraw = 1; Chartmaster = 0), *Order* (Shown First = 1; Shown Second = 0) and *Video Seen* (yes = 1, no = 0) are all binary dummy variables. The following regression equation was analyzed using ordinary least squares (OLS):

$$\text{Response} = b_1 + b_2\,\text{System} + b_3\,\text{Order} + b_4\,\text{Video Seen} + e$$

Table 7 shows the results of this regression equation applied to each of the four dependent variables. The R^2 value is a measure of the proportion of variability in the dependent (response) variable that is jointly explained by the independent (causal) variables. R^2 can range from 0.0 to 1.0, with higher values indicating a highly explanatory regression model. The beta (β) coefficients are standardized regression weights that serve as a measure of how much individual influence each independent variable has on the dependent variable. Beta coefficients can be interpreted as the number of units increase in the dependent variable resulting from a unit increase in the independent variable (where the variables are expressed in standardized units) while holding constant the other independent variables. Finally, the significance level is a measure of whether the magnitude of the effect of the independent variable (the beta coefficient) is significantly different from zero, with smaller values indicating greated significance. For an effect to be considered statistically significant, its

Table 7. Test for Equivalence of Hands-on Reactions With vs. Without having Previously Viewed Videotape

Dep. Var.	R^2	Independent Variable	b	S.E. (b)	β	t. stat.	sig. lvl.
EOU	.467	Constant	5.992	.222		27.035	.000
		System	-1.625	.222	-.614	-7.332	.000
		Order	.792	.222	.299	3.572	.001
		Video Seen	.058	.222	.022	.263	.793
USEF	.357	Constant	4.767	.295		16.151	.000
		System	-1.646	.295	-.513	-5.577	.000
		Order	.879	.295	.274	2.979	.004
		Video Seen	.478	.295	.136	1.482	.142
ATT	.326	Constant	5.344	.253		21.150	.000
		System	-1.313	.253	-.489	-5.195	.000
		Order	.775	.253	.289	3.067	.003
		Video Seen	.150	.253	.056	.594	.555
BE	.327	Constant	4.875	.348		14.006	.000
		System	-1.838	.348	-.497	-5.279	.000
		Order	.963	.348	.260	2.765	.007
		Video Seen	.413	.348	.112	1.185	.240

Note: Based on the hands-on reactions of all groups (n = 80)

significance level must be below .05. This implies that the probability of incorrectly concluding that beta is different from zero when it is actually equal to zero is .05.

As the Table 7 results show, having previously seen the videotape has no discernible effect on the hand-on ratings (significance level > .10 in all four equations). System and Order, as hypothesized, are highly significant, (sig. lvl. <.01 in all four equations for both variables). The signs on the coefficients indicate that showing a system first as opposed to second gives it an advantage, and that Chartmaster is significantly better than Pendraw across all response variables. Although not reported here, a separate regression analysis including all interaction effects showed that having seen the videotape did not interact significantly with any of the other variables in the regression. Thus, we regard the hands-on reactions of Groups 1 & 2 to be not substantially contaminated by carry-over effects, and pool these responses together with those of Groups 3 & 4 for subsequent analyses, thus yielding a total of 120 response observations (20 subjects x 2 systems = 40 for video plus 40 subjects x 2 systems = 80 for hands on) to be analyzed for testing the difference between video and hands-on based ratings.

We are now ready to test the effect of Medium (videotape = 1; hands-on = 0) on response ratings. Again we include System and Order in the regression to control for their effects on responses, which we now know are significant. The following main effects regression equation is used:

$$\text{Response} = b_1 + b_2\,\text{System} + b_3\,\text{Medium} + b_4\,\text{Order} + e$$

Table 8 shows the results of this regression analysis. Across all four response variables, the effect of Medium is nonsignificant. This is consistent with hypotheses regarding USEF, ATT and BE, but inconsistent with our expected effect of medium on EOU. An additional analysis of the above regression plus all interaction effects was carried out and is reported in Tables 9 (EOU & USEF) and 10 (ATT and BE). A summary of significant effects for the main effects and main + interaction effects regressions is presented in Table 11.

Table 8. Results of Main Effects Regression by Variable

Dep. Var.	R^2	Independent Variable	b	S.E. (b)	β	t. stat.	sig. lvl.
EOU	.240	Constant	5.751	.183		31.479	.000
		System	-.964	.195	-.400	-4.935	.000
		Medium	.142	.207	.055	.684	.495
		Order	.669	.195	.277	3.427	.001
USEF	.256	Constant	4.969	.229		21.735	.000
		System	-1.394	.244	-.457	-5.706	.000
		Medium	-.006	.259	-.002	-.024	.981
		Order	.66	.244	.22	2.71	.008
ATT	.292	Constant	5.392	.185		29.116	.000
		System	-1.154	.198	-.456	-5.830	.000
		Medium	.319	.210	.119	1.518	.132
		Order	.671	.198	.265	3.389	.000
BE	.255	Constant	5.085	.262		19.407	.000
		System	-1.608	.280	-.460	-5.741	.000
		Medium	-.019	.297	-.005	-.063	.950
		Order	.725	.280	.207	2.588	.011

For EOU, the main & interaction effects regression analysis found both a main effect for medium (whereas no main effect was found in the main effects only regression), and a very significant interaction between medium and system. The interaction effect implies that the effect of medium on EOU responses is different for Pendraw and Chartmaster. Table 12 shows the differential effect of medium on the two different systems that underlies the interaction effect for EOU. Having seen the videotape, participants formed unrealistically high EOU

Table 9. Results of Main Plus Interaction Effects Regression - EOU & USEF

Dep. Var.	R^2	Independent Variable	b	S.E. (b)	β	t. stat.	sig. lvl.
EOU	.426	Constant	6.267	.212		29.615	.000
		System	-2.117	.299	-.877	-7.073	.000
		Medium	-.800	.367	-.313	-2.183	.031
		Order	.300	.299	.124	1.003	.318
		S X M	2.250	.518	.695	4.341	.000
		S X O	.983	.423	.353	2.324	.022
		M X O	-.100	.518	-.031	-.193	.847
		S X M X O	-.533	.733	-.122	-.728	.468
USEF	.363	Constant	5.358	.282		19.02	.000
		System	-2.392	.398	-.784	-6.003	.000
		Medium	.125	.488	.039	.256	.798
		Order	.133	.398	.044	.335	.739
		S X M	.392	.690	.096	.568	.572
		S X O	1.492	.563	.423	2.647	.009
		M X O	-1.017	.690	-.248	-1.473	.144
		S X M X O	.725	.976	.131	.743	.459

perceptions for Pendraw, and unrealistically low EOU perceptions for Chartmaster, relative to hands-on based perceptions. These differential effects tend to offset each other, acting to mask the main effect of medium on EOU. Notice on Table 12 that similar, although non-significant, patterns were observed for the other response variables as well. Thus, our hypothesis of a significant effect of medium on EOU response is upheld with the qualification that the effect is highly system-dependent.

Table 10. Results of Main Plus Interaction Effects Regression - ATT & BE

Dep. Var.	R^2	Independent Variable	b	S.E. (b)	β	t. stat.	sig. lvl.
ATT	.321	Constant	5.600	.242		23.177	.000
		System	-1.675	.342	-.661	-.490	.000
		Medium	.200	.419	.074	.478	.634
		Order	.413	.342	.163	1.207	.230
		S X M	.550	.592	.162	.929	.355
		S X O	.725	.483	.248	1.500	.136
		M X O	-.238	.592	-.070	-.401	.689
		S X M X O	-.150	.837	-.033	-.179	.858
BE	.336	Constant	5.500	.330		16.681	.000
		System	-2.675	.466	-.765	-5.737	.000
		Medium	.050	.571	.013	.088	.930
		Order	.125	.466	.036	.268	.789
		S X M	.575	.808	.123	.712	.478
		S X O	1.675	.659	.415	2.540	.013
		M X O	-.825	.808	-.176	-1.022	.309
		S X M X O	.225	1.142	.036	.197	.844

Returning our attention to the regression results reported in tables 8-10 and summarized in Table 11, we observe that the significant main effect of order disappears in the main & interaction effects regression, being replaced by a significant System-Order Interaction for all response variables except for ATT, where this interaction effect only approached significance ($p = .136$). This says that the effect of order on response is highly system-specific, which is illustrated in Table 13. For Pendraw, a large drop in value was observed when shown as the second system as compared to being shown first. In contrast, the differences

Table 11. Summary of Significant Effects of System, Medium and Order on System Ratings

Regression Equation	Independent Variable	Dependent Variable			
		EOU	USEF	ATT	BE
Main effects	System	***	***	***	***
	Medium				
	Order	***	**	***	*
Main & Inter-	System	***	***	***	***
action effects	Medium	*			
	Order				
	S X M	***			
	S X O	*	*		*
	M X O				
	S X M X O				

Note: *P<.05; **p< .01; ***p<.001

Table 12. Differential Effect of Medium by System

System	Variable	Video (Groups 1& 2)	Hands-on (all groups)	Difference
Pendraw	EOU	5.93	4.79	1.14
	USEF	4.15	3.78	.37
	ATT	5.05	4.49	.56
	BE	4.05	3.73	.32
Chartmaster	EOU	5.57	6.42	-.85
	USEF	5.04	5.43	-.39
	ATT	5.89	5.81	.08
	BE	5.20	5.56	-.36

Table 13. Differential Effect of Order by System

System	Variable	Shown First	Shown Second	Difference
Pendraw	EOU	5.71	4.63	1.08
	USEF	4.67	3.14	1.53
	ATT	5.18	4.18	1.00
	BE	4.63	3.03	1.60
Chartmaster	EOU	6.27	6.00	.27
	USEF	5.19	5.40	-.21
	ATT	6.00	5.67	.33
	BE	5.37	5.52	-.15

for Chartmaster were generally much smaller in magnitude, with the responses for the USEF and BE variables even improving in value for Chartmaster when shown second.

Summarizing the above regression analyses, (1) Chartmaster is rated as superior to Pendraw across all response variables when medium and order are controlled for; (2) medium has a significant effect on EOU perceptions, this effect being highly system specific with video favoring Pendraw and hands-on favoring Chartmaster; and (3) there is a positive bias favoring whichever system is shown first, this effect being much greater for Pendraw than Chartmaster.

An alternative technique for assessing whether video serves as a good substitute for hands-on is to ask "to what extent is the video rating able to predict the hands-on rating for the same person and system". This is addressed using the following regression:

Hands-on Rating = a + b (Video Rating) + e

This analysis was performed for all of the basic TAM variables as well as for perceived quality of the output. It was not possible to include anticipated enjoyment of using in this analysis since no video-based measurement of this variable were taken. The results are reported in Table 14. Video ratings are

Table 14. Prediction of Hands-on Responses from Video Responses

Dep. Var.	R^2	Independent Variable	b	S.E. (b)	β	t. stat.	sig. lvl.
HEOU	.011	Constant	4.761	1.342		3.549	.001
		VEOU	.152	.231	.106	.658	.514
HUSEF	.463	Constant	.814	.730		1.116	.272
		VUSEF	.872	.152	.680	5.722	.000
HATT	.553	Constant	-.133	.797		-.167	.868
		VATT	.980	.143	.743	6.851	.000
HBE	.400	Constant	1.202	.764		1.574	.124
		VBE	.789	.157	.632	5.028	.000
HQUAL	.126	Constant	2.529	1.068		2.368	.023
		VQUAL	.455	.195	.355	2.339	.025

significantly related to hands-on ratings for all of the variables except for Ease of Use, which agrees with the findings reported above.

Confidence

As discussed previously, a person's confidence in their ratings is a potentially important diagnostic metric for video based ratings. Table 15 presents a

Table 15. Confidence of Ratings in Experiment

Variable	Mean Confidence Ratings			Significance Level of Difference		
	A Video Groups 1&2	B Hands-on Groups 1&2	C Hands-on Groups 3&4	A vs. C	A vs. B	B vs. C
EOU	6.20	6.53	5.95	.312	.161	.018
USEF	5.90	6.38	5.88	.920	.071	.028
ATT	5.83	6.55	6.00	.452	.001	.004
BE	5.78	6.33	5.98	.319	.015	.087
FUN	NA	6.63	6.50	NA	NA	.348
QUAL	5.85	6.55	6.20	.139	.003	.014

comparison of confidence ratings between Group 1 and 2 video responses, Group 1 and 2 hands-on responses and Group 3 and 4 hands-on responses. A comparison of these subgroups shows that there is no significant difference between video ratings for group 1 and 2 and the hands-on ratings of group 3 and 4 for any of the variables. This is counter to Fazio and Zanna's (1981) theory that attitudes based on direct experience should be more confidently held. One possible interpretation is that the single-item measurement of confidence is not reliable enough to detect a true difference. The hands-on based responses of group 1 and 2 are generally significantly greater than both the group 1 and 2 video responses and the group 3 and 4 hands-on responses. A possible interpretation of this is that greater exposure time leads to greater attitude confidence (since group 1 and 2 hands-on respondents received both a video and a hands-on exposure to the systems). An additional way of analyzing the role of confidence is to see if the video ratings of people who are more confident in their ratings predict their own hands-on ratings better than the ratings of those who are less confident. Operationally, this would be shown if the video ratings and the confidence in those ratings exhibited a significant interaction effect in predicting hands-on ratings. Table 5A.1 (see Appendix) contains the results of regression analyses of this. We observe that the interaction terms has a large beta coefficients associated with them for all dependent variables. This suggests that an interaction term is present. However, these coefficients failed to achieve statistical significane due to their very high standard errors which are likely a consequence of the naturally high collinearity beetween the interaction term and the main effects. In fact the EOU regression could not be analyzed in this way due to extremely high collinearity between the interaction term and the main effects. We may conclude that rating confidence has a potential role in diagnosing the trustworthiness of video based measures, but more research will be needed before we can fully understand the role it plays.

In this section, a series of regression analyses performed for the purpose of validating the causal structure of Technology Acceptance Model are reported. Only the hands-on responses were used for this analysis, yielding 80 observations (40 subjects x 2 systems). Table 16 contains results of four

Table 16. TAM Model Test Regressions: Tests of Significant Relationships

Dep. Var.	R^2	Independent Variable	b	S.E. (b)	β	t. stat.	sig. lvl.
BE	.741	Constant	-.647	.445		-1.454	.150
		ATT	.462	.177	.335	2.615	.011
		USEF	.632	.148	.549	4.276	.000
ATT	.818	Constant	1.089	.287		3.792	.000
		USEF	.664	.049	.794	13.555	.000
		EOU	.180	.059	.177	3.030	.003
USEF	.385	Constant	2.815	.933		3.018	.004
		EOU	.360	.149	.297	2.415	.018
		System	-1.060	.377	-.330	-2.813	.006
		Order	.594	.312	.185	1.905	.061
EOU	.467	Constant	6.021	.191		31.560	.000
		System	-1.625	.220	-.614	-7.377	.000
		Order	.792	.220	.299	3.594	.001

regression analyses that were used to estimate the hypothesized significant relationships of the Model. The first regression estimates the effect of ATT and USEF on BE. The significant effect of USEF indicates that it has a causal effect on BE over and above that of ATT alone, which confirms findings from the survey reported in Chapter 4. The second equation assesses the effect of USEF and EOU on ATT. The third regression estimates the effect of EOU, System, and Order on USEF, with all effects found to be significant. Finally, the fourth regression exhibits the significant effects of Order and System on EOU.

The percent of variance explained (R^2) values for the USEF and EOU regressions are lower than those found in the other two regressions of Table 16. This suggests that different people exhibit varying EOU and USEF perceptions regarding the same system. There are many reasons why this makes sense. People will have differing job needs that will influence their perceived usefulness of the same system. People may have different experience and ability levels that may cause differences in how easy to use a given system is perceived to be. Finally, differing individual styles in responding to questionaire rating scales may also produce differences.

Table 17 gives the regressions performed for the purpose of demonstrating that relationships hypothesized to be non-significant are actually non-significant. F-tests which compare the R^2 values for restricted and unrestricted regressions are used to test whether adding the additional independent variables into the regression explain a significantly greater amount of the variance in the dependent variable. The first equation establishes that adding system, order and EOU does not increase the explained variance of BE ($F(3,74) = .991$, n.s.). This confirms the finding from the survey (Davis, 1985) that EOU does not have any direct effect on behavior. In addition, whereas in the survey system had a significant effect on attitude over and above perceived usefulness and perceived ease of use (as discussed later), this was not observed in the experiment. The second equation shows that system and order do not add explanatory ability above and beyond EOU and USEF ($F(2,75) = .208$, n.s.). Thus System and Order have no direct effects on ATT or BE, and must therefore

Table 17. Model Test Regressions: Tests of Insignificant Relationships

Dep. Var.	R^2	Independent Variable	b	S.E. (b)	β	t. stat.	sig. lvl.
BE	.751	Constant	-.931	.752		-1.238	.220
		ATT	.362	.188	.263	1.930	.058
		USEF	.620	.152	.538	4.075	.000
		EOU	.159	.121	.114	1.319	.191
		System	-.083	.295	-.022	-.282	.779
		Order	.011	.237	.003	.045	.964
ATT	.819	Constant	.882	.451		1.955	.054
		USEF	.671	.052	.802	12.801	.000
		EOU	.198	.071	.195	2.797	.007
		System	.113	.181	.042	.627	.533
		Order	.028	.146	.011	.195	.846

exert their causal influence entirely indirectly via EOU and USEF. This is consistent with the Technology Acceptance Model and the theory upon which it is founded (e.g., Fishbein & Ajzen, 1975).

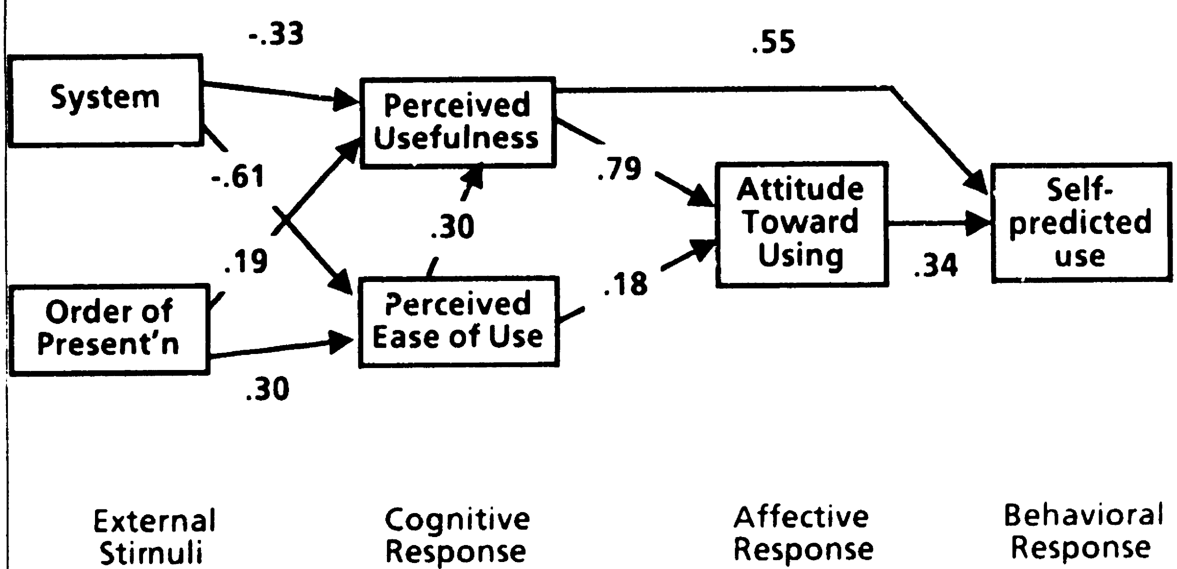
The regressions were used to estimate the causal parameters of the model. Table 18. gives the point and 95% confidence interval estimates for model parameters. The 95% confidence intervals specify a range such that the likelihood that the true parameter for which the beta coefficient provides an estimate falls inside the range is .95. The estimated beta coefficients lead to the causal diagram given in Figure 4. Perceived usefulness has a stronger influence on self-predicted use (.55) than does attitude toward using (.34). Perceived usefulness has a much stronger influence on attitude toward using than does

Table 18. TAM Parameter Estimates and 95% Confidence Intervals-Experiment Data

Causal Link		Point Estimate			95% Confidence Interval	
Ind. Var.	Dep. Var.	β	Std. Error	Sig. Level	Lower Bound	Upper Bound
System	EOU	-.614	.083	.000	-.779	-.449
Order	EOU	.299	.083	.001	.134	.464
System	USEF	-.330	.117	.006	-.563	-.097
Order	USEF	.185	.097	.061	-.008	.378
EOU	USEF	.297	.123	.018	.052	.542
EOU	ATT	.177	.058	.003	.062	.292
USEF	ATT	.794	.059	.000	.677	.911
USEF	BE	.549	.129	.000	.292	.806
ATT	BE	.335	.128	.011	.080	.590

perceived ease of use (.79 vs. .18). Perceived ease of use has a moderate effect on usefulness (.30). The negative coefficients for the system-perception relationship imply a more negative rating for Pendraw than Chartmaster (this is due to the fact that the dummy variable for system was coded as 1 for Pendraw and 0 for Chartmaster, resulting in a negative relationship between the system variable and perceptions. It should be pointed out that this choice of variable coding was arbitrary, and the reverse coding of the variable would reverse the sign of the coefficient, but not change its magnitude nor its interpretation). The direct effect of system on ease of use was greater than its effect on usefulness (-.61 vs. -.33). Finally, order of presentation had a significant effect on both usefulness (.19) and ease of use (.30).

Figure 4. Causal Diagram of Model Validation Results



Note: Regression beta coefficients are shown

The causal diagram enables one to analyze the relative importance of improving EOU as compared to USEF. Increasing USEF by one unit will exert a direct effect on BE of .55 plus an indirect effect via ATT of .79 x .34. This adds up to .82 units increase in BE. EOU has an effect through ATT of .18 x .34 plus an effect through USEF of .30 x .82 (where the .82 is USEF's effect on BE both direct and indirect). This totals .31. Therefore, USEF is about 2.65 times as important as EOU in influencing BE based on this experiment. Our ability to generalize this finding is limited, and awaits further data that will give us a perspective on how much this value may vary from one context to another. For example, the survey data (reported in Chapter 4) found that USEF was only 1.49 times as important as EOU in influencing USE. A key difference in the survey is that EOU had a much greater influence on USEF than was found in the experiment (.64 vs. .30).

Evaluation of Pendraw (r) and Chartmaster (r)

Table 20 gives a comparison of the mean values between the two systems evaluated, Chartmaster and Pendraw, for each of the variables measured in the experiment. These means are based on hands-on ratings only. For all of the variables except EOU_{non} and $QUAL_{non}$, the ratings for Chartmaster are significantly greater than those for Pendraw. As hypothesized, the non-numeric task-specific ratings for EOU and QUAL favor Pendraw. However, this difference is not statistically significant. One reason for the lack of significance relates to the potential problems identified in the task-specific scales discussed in a later section. Also, as will be addressed below, the reasons why the favorability of Pendraw for non-numeric tasks did not translate into overall favorability probably relate to the relative importance of numeric graphs ($IMPORT_{num}$ = 5.6) as compared to non-numeric graphs ($IMPORT_{non}$ = 4.8) for this user audience.

Table 20. Comparison of Mean Ratings by System

Variable	CHARTMASTER		PENDRAW		Difference		
	Mean	Std. Dev.	Mean	Std. Dev	Mean Diff.	t-stat.	Sig.Lvl.
EOU	6.417	.505	4.792	1.408	1.625	6.87	.000
USEF	5.425	.970	3.779	1.717	1.646	5.28	.000
ATT	5.806	.898	4.494	1.415	1.312	4.95	.000
BE	5.563	1.215	3.725	1.951	1.838	5.06	.000
QUAL	5.708	1.060	4.392	1.511	1.316	4.51	.000
FUN	6.025	.907	5.308	1.611	.717	2.45	.016
EOU_{num}	6.100	.709	3.050	1.319	3.050	12.88	.000
EOU_{non}	3.564	2.337	4.450	1.568	-.886	-1.98	.051
$QUAL_{num}$	6.375	.838	3.925	1.73	2.450	8.06	.000
$QUAL_{non}$	4.026	2.497	4.950	1.518	-.924	-1.99	.050

Note: Hands-on ratings only (n = 80), averaged over order of presentation. Ratings may range from 1 (low) to 7 (high).

Theoretical Elaborations of TAM

TAM2

Tables 5A.2 and 5A.3 give the restricted and unrestricted regressions for testing the proposed TAM2 model presented earlier. T-tests on Table 5A.3 demonstrate the non-significance of the direct effects from: QUAL to FUN and from FUN to USEF. F-tests verify that: System, EOU, and QUAL do not have significant direct effects on ATT above and beyond FUN and USEF (F (3, 74) = 2.00, n.s.); and all other variables affect BE only indirectly via their effects on ATT and USEF (F (4, 74) = 1.72, n.s.). The hypothesis that EOU does not effect QUAL was resoundingly rejected however (see Table 5A.3). This was quite a surprising result, since it was theorized that EOU and QUAL would be causally unrelated constructs. Upon retrospection, it was reasoned that in the case of Pendraw, the quality of the graph is very sensitive to the effort and ability of the user. Thus people who found Pendraw easier to use would be likely to make graphs of a higher quality. This would be much less the case for Chartmaster, where the quality of the graph that is produced is comparatively invariant to the actions of the user. If this reasoning is correct, then when the EOU-QUAL link is examined separately for each system, we should observe a strong link for Pendraw, but not for Chartmaster. This is exactly what was found (see table 5A.4). Thus, at least for certain classes of systems, there is a theoretically meaningful causal relationship from EOU to QUAL. This link was therefore included as a revision to TAM2.

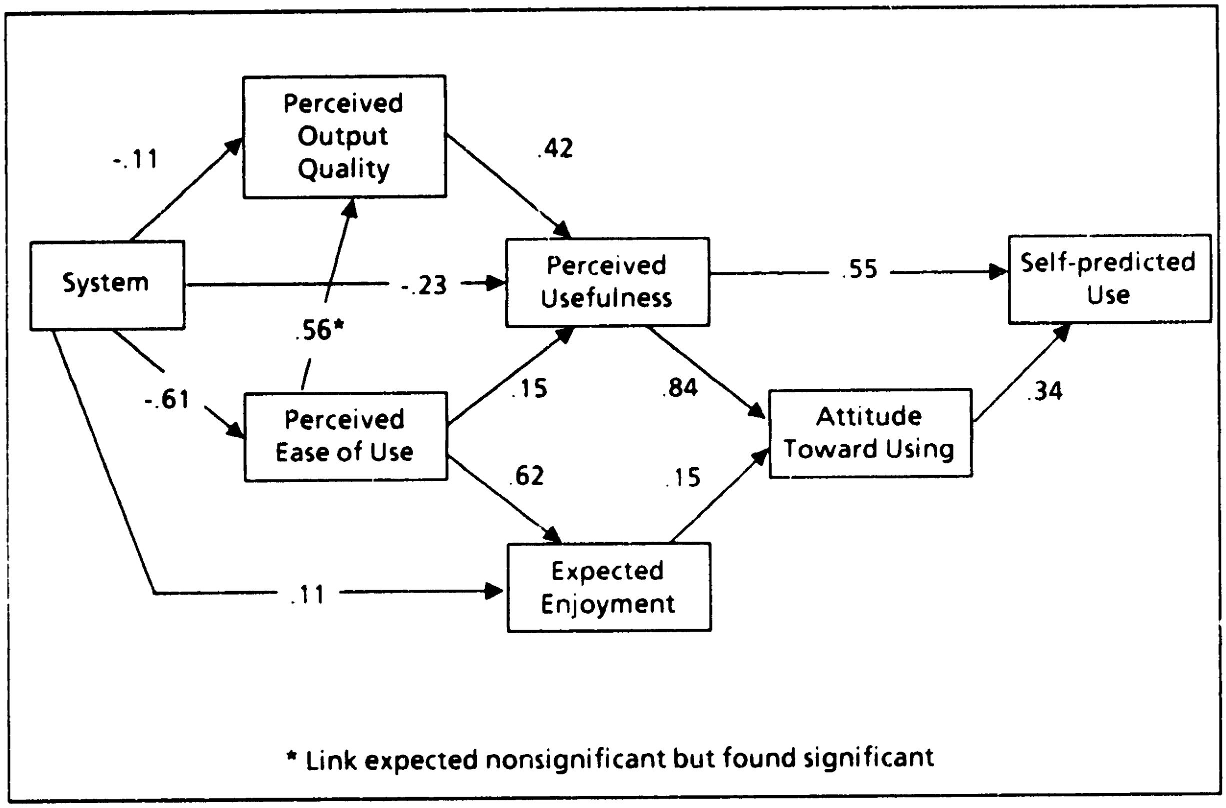
The estimated coefficients based on the revised model are given in Table 21. The parameter estimates are also specified and the diagram in Figure 6. Notice that the effect of system on QUAL operates largely through EOU. EOU has a strong influence on enjoyment, although enjoyment is only weakly linked to ATT. QUAL turned out to be a rather important variable, having a farily strong influence on USEF, which has been shown to be a powerful determinant of user behavior. Notice that some of the links hypothesized to be significant were found to be insignificant. These are left in the model nevertheless for several reasons. They were specified based on theoretical arguments. To assume they are negligible because of insignificant statistical tests runs into the well-known problem of accepting the null hypothesis. There are a number of reasons why their magnitude may be insignificant even though there exist a true underlying relationship: restriction of range of the variables, insufficient powere of the tests, inflated standard error due to collinearity. Further, to insure consistent estimation of the other parameters of the model, these nonsignificant

Table 21. TAM2 Parameter Estimates and 95% Confidence Intervals

Causal Link		Point Estimate			95% Confidence Interval	
Ind. Var.	Dep. Var.	β	Std. Error	Sig. Level	Lower Bound	Upper Bound
System	EOU	-.614	.089	.000	-.792	-.436
System	QUAL	-.113	.112	.316	-.336	.110
EOU	QUAL	.557	.112	.000	.335	.779
System	FUN	.113	.120	.348	-.125	.351
EOU	FUN	.620	.120	.000	.382	.858
System	USEF	-.229	.107	.037	-.443	-.015
EOU	USEF	.153	.123	.217	-.092	.398
QUAL	USEF	.419	.108	.000	.204	.634
FUN	ATT	.149	.053	.006	.044	.254
USEF	ATT	.836	.053	.000	.731	.941
USEF	BE	.549	.129	.000	.293	.805
ATT	BE	.335	.128	.011	.080	.590

relationships should remain in the model.

Figure 6. TAM2 Parameter Estimates



TAM3

Tables 5A.6, 5A.7 and 5A.8 give the constrained and unconstrained regressions for TAM3 (without incorporating importance weights for the time being). T-tests on Table 5A.7 show that the EOU_{num}-EOU_{non}, EOU_{non}-$QUAL_{num}$, and FUN-USEF relationships are non-significant, as hypothesized. F-tests verify that no variables above and beyond FUN and USEF have a direct influence on ATT (F (5, 71) = .48, n.s.) and that no variables other than ATT and USEF directly influence BE (F (6,70) = 2.097, n.s.). However, the hypothesized non-significance of the effects of EOU_{num} and $QUAL_{num}$ on $QUAL_{non}$ was disconfirmed ($F(2, 74) = 12.39, p < .01$). This result was unexpected since these effects cross over from one task domain (numeric charts) to another (non-numeric charts). One possible interpretation in terms of the judgemental processes by which users may form QUAL and EOU perceptions is addressed below. For the time being, however, these non-hypothesized links were included in the estimation of parameters so as to reduce the possible biasing effects of "omitted varibles".

Table 22 gives the point estimates and confidence intervals estimated for TAM 3. These are also specified on the TAM3 causal diagram, Figure 7. Notice that the effects of numeric perceptions on USEF and FUN are greater than their non-numeric counterparts. This is consistent with our observation of a greater mean importance for numeric graphs ($IMPORT_{num}$ mean = 5.6, std dev. = 1.2; $IMPORT_{non}$ mean = 4.8, std. dev. = 1.5). We should bear in mind, however, that the mean values for the importance weights may have been inflated if the experimental procedure made numeric graphing more salient by placing greater emphasis on it. Referring again to Figure 7, notice that the effects of system on QUAL appear to be entirely indirect via EOU. Again, this may be due to the nature of Pendraw and the resulting impact of EOU on QUAL, as discussed in the context of TAM2 above. The direct effect of system on usefulness has now diminished to a non-significant level.

We will now examine whether the relationships between task-specific perceptions and USEF is moderated by subjects' task importance ratings, as hypothesized. The appropriate regressions are given in Table 5A.9. The beta coefficients for the interaction term varied in magnitude from .26 to .64, although none were statistically significant. This does not necessarily imply that importance does not moderate the link between task-specific perceptions and

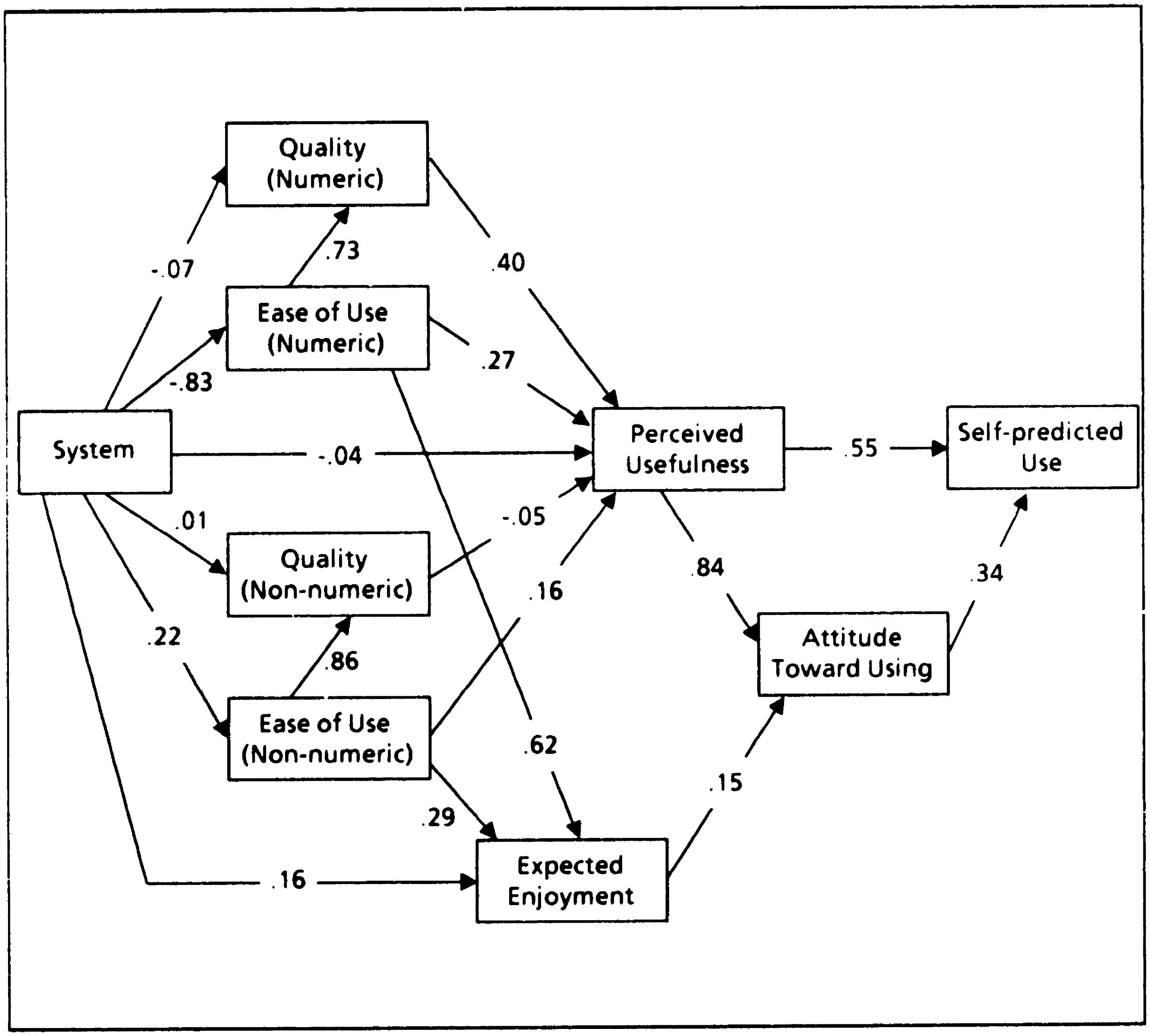
Table 22. TAM3 Parameter Estimates and 95% Confidence Intervals

Causal Link		Point Estimate			95% Confidence Interval	
Ind. Var.	Dep. Var.	β	Std. Error	Sig. Level	Lower Bound	Upper Bound
System	EOU_{num}	-.825	.064	.000	-.953	-.697
System	EOU_{non}	.220	.111	.051	-.001	.441
System	$QUAL_{num}$	-.073	.555	.123	-.318	.172
EOU_{num}	$QUAL_{num}$	.729	.123	.000	.484	.974
System	$QUAL_{non}$	.010	.088	.910	-.165	.185
EOU_{num}	$QUAL_{non}$	-.350	.102	.001	-.553	-.147
EOU_{non}	$QUAL_{non}$	.861	.050	.000	.761	.961
$QUAL_{num}$	$QUAL_{non}$	.396	.078	.000	.241	.551
System	FUN	.162	.175	.359	-.187	.511
EOU_{num}	FUN	.615	.171	.001	.274	.956
EOU_{non}	FUN	.294	.101	.004	.094	.494
System	USEF	-.041	.162	.800	-.364	.282
EOU_{num}	USEF	.271	.201	.181	-.128	.670
EOU_{non}	USEF	.160	.204	.434	-.245	.565
$QUAL_{num}$	USEF	.399	.164	.018	.072	.726
$QUAL_{non}$	USEF	-.052	.212	.806	-.474	.370

USEF, since the range of the importance variable in the population studied may have been restricted. Moreover, the crude single item rating scales used for measuring task-specific perceptions are unvalidated and their unreliability may act to attenuate observed relationships.

The pattern of responses for the task-specific rating scales was examined in an effort to gain some insight regarding what may account for the unexpected

Figure 7. TAM3 Parameter Estimates



effects of EOU_{num} and $QUAL_{num}$ on $QUAL_{non}$ (see Table 5A.10). One of the characteristics of these scales is that they attempt to incorporate the infeasible case as one of their endpoints. For example, the difficult end of the EOU scale has "impossible" as its anchoring adjective. Similarly, QUAL has "unacceptable (or non-existent)" at the low end of the scale. Thus, when asked a task-specific question regarding a system that is technically incapable of performing that task, the respondent should logically choose the negative endpoint (assuming they understand that the system is incapable of performing the task). This would be the case when asking subjects non-numeric perception for Chartmaster. The pattern of responses (Table 5A.10) shows that while more than 25% of the subjects selected the lowest scale point, more than 2/3 selected something else. In fact, the answers were quite distributed across the scale points. A couple of possible interpretations of non-low-endpoint reponses may be made. First, the subject may have correctly realized that Chartmaster was incapable of non-numeric charts but did not properly understand the rating scale. Alternately, the subject may not have realized that Chartmaster was in fact restricted to numeric charts. Given the large number of people responding in this manner, coupled with the fact that Chartmaster's inability to handle non-numeric graphs was not heavily emphasized in the experimental procedure, it appears that the latter interpretation is likely to be appropriate for many of the subjects. Faced with uncertainty about whether or not Chartmaster is able to handle non-numeric charts and about the quality and/or ease of use of making non-numeric charts with Chartmaster, subjects may base their non-numeric perceptions on their numeric perceptions. Since the major experimental excercise for Pendraw was to create a numeric chart (although subjects made partial non-numeric charts when learning the systems features), a similar process may take place there as well. This is one possible explanation about why there is such a strong influence observed from EOU_{num} and $QUAL_{num}$ on $QUAL_{non}$. The corresponding effect of EOU_{num} on EOU_{non} was nearly significant (sig. level = .06).

Discussion

These findings suggest a promising role for videotape presentation as a substitute for hands-on interaction in applied user acceptance testing contexts. Videotape allows accurate formation of those perceptions and attitudes that most strongly influence user behavior. The inaccuracy of videotape for conveying ease of use is mitigated by the fact that ease of use appears to be a less powerful determinant of system acceptance. Thus, although videotape is not an advisable medium for studying ease of use issues in systems design, it does appear to be a suitable medium for addressing issues relating to the functional content of the system, which appear to be more critical in the ultimate success of the system.

It is too early to attempt to generalize these findings regarding videotapes to all possible videotape applications. We should attempt to gain an understanding of the critical characteristics of videotapes that affect their ability to substitute for hands-on interaction. One obvious characteristic is the quality of the videotapes. The quality of the videotapes used in this experiment was quite high in terms of content and presentation style. It would no doubt be erroneous to conclude that these results would be applicable to videotapes of significantly reduced quality. The substitutability of videotape for hands-on has been addressed for the first time in this experiment. To generalize these results to other contexts is premature at this time, and awaits additional evidence that may accumulate as video-based user acceptance testing begins to be applied in practice.

The differential bias of video-tape based ease of use perceptions for the two systems is an interesting finding that suggests that the nature of the system's interface may be an important determinant of the ability to convey ease of use by video. Pendraw has a highly graphical and highly direct user interface which may be easier to present in a videotape than Chartmaster's menu-driven interface. The apparent relative simplicity of Pendraw's interface as conveyed by video was later reversed in the hands-on demonstration, which may have suggested that Pendraw's interface required greater motor dexterity than Chartmaster's.

Regarding the significantly superior ratings of Chartmaster relative to Pendraw, it is difficult to attribute this difference to any single system characteristic. The systems differed across numerous features and characteristics. The only way to isolate any particular characteristic as being significant would be to compare systems against each other where only the single feature of interest differs between them.

The results of the technology acceptance model (TAM) validation largely agree with the findings of the survey reported in Chapter 4, which increases our confidence in these results. As evidence gradually accumulates across a number of contexts, we begin to gain an understanding of the extent to which the magnitudes of the causal relationships being modeled vary across contexts. In this way, the external validity of those findings is established, which means we may begin to confidently generalize those findings to contexts which have not previously been explicitly tested. As part of this iterative process, it is appropriate to explore extensions, refinements and elaborations of the causal model. This will tend to lead toward a more comprehensive understanding of user acceptance processes, and enable us to devise increasingly powerful system design and evaluation methodologies.

Table 5A.1. Regression Tests of Moderating Effect of Confidence on Predicting Hands-on Ratings from Video Ratings

Dep. Var.	R^2	Independent Variable	b	S.E. (b)	β	t. stat.	sig. lvl.
HUSEF	.488	Constant	-1.877	4.055		-.463	.646
		VUSEF	1.797	1.000	1.402	1.796	.081
		VUSEFC	.385	.633	.275	.608	.547
		Interaction	-.139	.154	-.878	-.903	.372
HATT	.603	Constant	-5.311	4.371		-1.215	.232
		VATT	1.713	.871	1.300	1.968	.057
		VATTC	.938	.717	.726	1.308	.199
		Interaction	-.133	.141	-.920	-.949	.349
HBE	.646	Constant	5.131	5.218		.983	.332
		VBE	-.318	1.197	-.255	-.266	.792
		VBEC	-.650	.882	-.329	-.737	.466
		Interaction	.183	.199	1.034	.920	.364
HQUAL	.136	Constant	-.741	5.153		-.144	.886
		VQUAL	1.025	.940	.799	1.090	.283
		VQUALC	.566	.868	.486	.652	.519
		Interaction	-.098	.155	-.722	-.630	.533

Table 5A.2. TAM2 Regression Tests of Proposed Model

Dep. Var.	R^2	Independent Variable	b	S.E. (b)	β	t. stat.	sig. lvl.
EOU	.377	Constant	6.417	.167		38.368	.000
		System	-1.625	.237	-.614	-6.871	.000
QUAL	.207	Constant	5.707	.206		27.658	.000
		System	-1.317	.292	-.455	-4.511	.000
FUN	.311	Constant	1.998	.799		2.501	.015
		System	.303	.321	.113	.944	.348
		EOU	.628	.121	.620	5 172	.000
USEF	.461	Constant	1.586	.878		1.805	.075
		System	-.733	.344	-.229	-2.129	.037
		EOU	.185	.149	.153	1.245	.217
		QUAL	.465	.120	.419	3.856	.000
ATT	.815	Constant	1.090	.300		3.635	.001
		FUN	.149	.053	.149	2.804	.006
		USEF	.699	.044	.836	15.773	.000
BE	.741	Constant	-.647	.445		-1.454	.150
		USEF	.632	.148	.549	4.276	.000
		ATT	.462	.177	.335	2.615	.011

Table 5A.3. TAM2 Regression Tests of Insignificant Relationships

Dep. Var.	R^2	Independent Variable	b	S.E. (b)	β	t. stat.	sig. lvl.
QUAL	.400	Constant	1.798	.805		2.233	.029
		System	-.326	.324	-.113	-1.008	.316
		EOU	.609	.122	.557	4.983	.000
FUN	.314	Constant	1.889	.828		2.281	.025
		System	.323	.325	.121	.994	.323
		EOU	.591	.140	.583	4.213	.000
		QUAL	.061	.114	.065	.533	.595
USEF	.470	Constant	1.336	.907		1.474	.145
		System	-.776	.346	-.241	-2.241	.028
		EOU	.107	.165	.088	.650	.518
		QUAL	.457	.121	.412	3.787	.000
		FUN	.132	.121	.110	1.085	.281
ATT	.827	Constant	.643	.443		1.453	.150
		System	.087	.172	.032	.503	.616
		EOU	.123	.080	.121	1.543	.127
		QUAL	.059	.063	.063	.928	.357
		FUN	.093	.059	.093	1.575	.119
		USEF	.641	.055	.767	11.54	.000
BE	.763	Constant	-1.232	.729		-1.690	.095
		System	-.119	.280	-.032	-.426	.671
		EOU	.039	.131	.028	.294	.769
		QUAL	.147	.103	.115	1.423	.159
		FUN	.125	.097	.091	1.285	.203
		USEF	.602	.151	.522	3.989	.002
		ATT	.291	.189	.211	1.543	.127

Table 5A.4. TAM2 Moderating Effect of System on EOU-QUAL Relationship

Dep. Var.	R^2	Independent Variable	b	S.E. (b)	β	t. stat.	sig. lvl.
QUAL	.400	Constant	1.798	.805		2.233	.029
(Pooled)		System	-.326	.324	-.113	-1.008	.316
		EOU	.609	.122	.557	4.983	.000
QUAL	.006	Constant	4.687	2.185		2.145	.038
(CHART)		EOU	.159	.340	.076	.469	.642
QUAL	.387	Constant	1.194	.680		1.755	.087
(PENDR)		EOU	.667	.136	.622	4.893	.000

Table 5A.5. TAM2 Intercorrelation Matrix

Variable	SYS	FUN	QUA	EOU	USF	ATT	BE
System							
FUN	-.27						
QUAL	-.46	.38					
EOU	-.61	.55	.63				
USEF	-.51	.38	.62	.56			
ATT	-.49	.46	.63	.62	.89		
BE	-.50	.45	.64	.59	.85	.82	

Table 5A.6. TAM3 Regression Tests of Proposed Model

Dep. Var.	R^2	Independent Variable	b	S.E. (b)	β	t. stat.	sig. lvl.
EOU_{num}	.680	Constant	6.100	.167		36.425	.000
		System	-3.050	.237	-.825	12.878	.000
EOU_{non}	.049	Constant	3.564	.318		11.211	.000
		System	.886	.447	.220	1.983	.051
$QUAL_{num}$	.624	Constant	2.007	.762		2.634	.010
		System	-.266	.449	-.073	.593	.555
		EOU_{num}	.716	.121	.729	5.898	.000
$QUAL_{non}$	.770	Constant	.804	.265		3.037	.003
		System	.124	.235	.030	.526	.601
		EOU_{non}	.904	.058	.871	15.456	.000
FUN	.325	Constant	2.642	.765		3.455	.001
		System	.434	.470	.162	.923	.359
		EOU_{num}	.445	.124	.615	3.593	.001
		EOU_{non}	.196	.067	.294	2.951	.004
USEF	.448	Constant	1.439	.877		1.641	.105
		System	-.131	.519	-.041	.254	.800
		EOU_{num}	.235	.174	.271	1.351	.181
		EOU_{non}	.128	.163	.160	.786	.434
		$QUAL_{num}$	.352	.145	.399	2.423	.018
		$QUAL_{non}$	-.040	.163	-.052	.246	.806
ATT	.815	Constant	1.090	.300		3.635	.001
		FUN	.149	.053	.149	2.804	.006
		USEF	.699	.044	.836	15.773	.000
BE	.741	Constant	-.647	.445		1.454	.150
		ATT	.462	.177	.335	2.615	.011
		USEF	.632	.148	.548	4.276	.000

Table 5A.7. TAM3 Regression Tests of Insignificant Relationships (Part A)

Dep. Var.	R^2	Independent Variable	b	S.E. (b)	β	t. stat.	sig. lvl.
EOU_{non}	.092	Constant	1.134	1.311		.864	.390
		System	2.102	.774	.523	2.716	.008
		EOU_{num}	.398	.209	.367	1.908	.060
$QUAL_{num}$	.625	Constant	1.971	.774		2.544	.013
		System	-.362	.476	-.099	-.760	.450
		EOU_{num}	.699	.126	.710	5.569	.000
		EOU_{non}	.041	.067	.054	.612	.542
$QUAL_{non}$	.831	Constant	.355	.624		.570	.571
		System	.042	.370	.010	.113	.910
		EOU_{num}	-.394	.115	-.350	-3.417	.001
		EOU_{non}	.894	.052	.861	17.120	.000
		$QUAL_{num}$	.453	.089	.396	5.070	.000
FUN	.330	Constant	2.569	.807		3.185	.002
		System	.457	.477	.170	.958	.341
		EOU_{num}	.373	.160	.515	2.331	.023
		EOU_{non}	.274	.150	.409	1.823	.072
		$QUAL_{num}$	.093	.134	.127	.697	.488
		$QUAL_{non}$	-.089	.150	-.138	-.592	.556
USEF	.453	Constant	1.179	.939		1.256	.213
		System	-.178	.523	-.055	-.340	.735
		EOU_{num}	.197	.181	.227	1.091	.279
		EOU_{non}	.101	.167	.126	.601	.550
		$QUAL_{num}$	.343	.146	.389	2.344	.022
		$QUAL_{non}$	-.031	.164	-.040	-.190	.850
		FUN	.101	.128	.085	.794	.430

Table 5A.8. TAM3 Regression Tests of Insignificant Relationships (Part B)

Dep. Var.	R^2	Independent Variable	b	S.E. (b)	β	t. stat.	sig. lvl.
ATT	.821	Constant	.936	.457		2.051	.044
		System	.111	.252	.041	.440	.661
		EOU_{num}	.021	.088	.029	.241	.811
		EOU_{non}	.089	.081	.133	1.107	.272
		$QUAL_{num}$	.078	.073	.106	1.075	.286
		$QUAL_{non}$	-.084	.079	-.131	-1.071	.288
		FUN	.117	.062	.116	1.889	.063
		USEF	.655	.057	.783	11.555	.000
BE	.780	Constant	-1.522	.724		-2.102	.039
		System	.164	.389	.044	.422	.675
		EOU_{num}	-.005	.135	-.005	-.036	.971
		EOU_{non}	.085	.125	.092	.680	.499
		$QUAL_{num}$	.266	.113	.262	2.345	.022
		$QUAL_{non}$	-.054	.123	-.061	-.441	.661
		FUN	.101	.097	.073	1.037	.303
		USEF	.574	.148	.499	3.874	.000
		ATT	.275	.183	.200	1.503	.137

Table 5A.9. TAM3 Regression Tests of Moderating Effects of Task Importance

Dep. Var.	R^2	Independent Variable	b	S.E. (b)	β	t. stat.	sig. lvl.
USEF	.481	Constant	1.316	1.482		.888	.378
		EOU_{num}	.216	.298	.249	.727	.470
		IMP_{num}	.147	.260	.110	.563	.575
		EOU*IMP	.057	.052	.423	1.088	.280
USEF	.081	Constant	3.667	1.446		2.535	.013
		EOU_{non}	-.099	.310	-.123	-.318	.752
		IMP_{non}	.142	.297	.129	.479	.633
		EOU*IMP	.036	.064	.256	.564	.574
USEF	.487	Constant	2.089	1.618		1.291	.201
		$QUAL_{num}$	.072	.312	.081	.230	.819
		IMP_{num}	-.051	.282	-.038	-.180	.857
		QUAL*IMP	.083	.054	.644	1.546	.126
USEF	.092	Constant	4.289	1.529		2.806	.006
		$QUAL_{non}$	-.225	.302	-.291	-.744	.459
		IMP_{non}	.000	.311	.000	.001	.999
		QUAL*IMP	.063	.061	.479	1.033	.305

Table 5A.10. Response Patterns of Task-specific Scales Used in TAM3

Scale Point	Frequency of Task-specific Ratings by Scale Point							
	EOU_{num}		$QUAL_{num}$		EOU_{non}		$QUAL_{non}$	
	CM*	PD*	CM	PD	CM	PD	CM	PD
1 (low)	0	2	0	3	13	0	11	0
2	0	16	0	8	5	7	4	4
3	0	7	0	4	2	4	3	1
4	1	11	2	10	2	8	3	12
5	5	2	3	7	4	9	1	6
6	23	1	13	5	9	9	7	10
7 (high)	11	1	22	3	4	3	10	7

* CM refers to CHARTMASTER
PD refers to PENDRAW

CHAPTER 6. GENERAL DISCUSSION

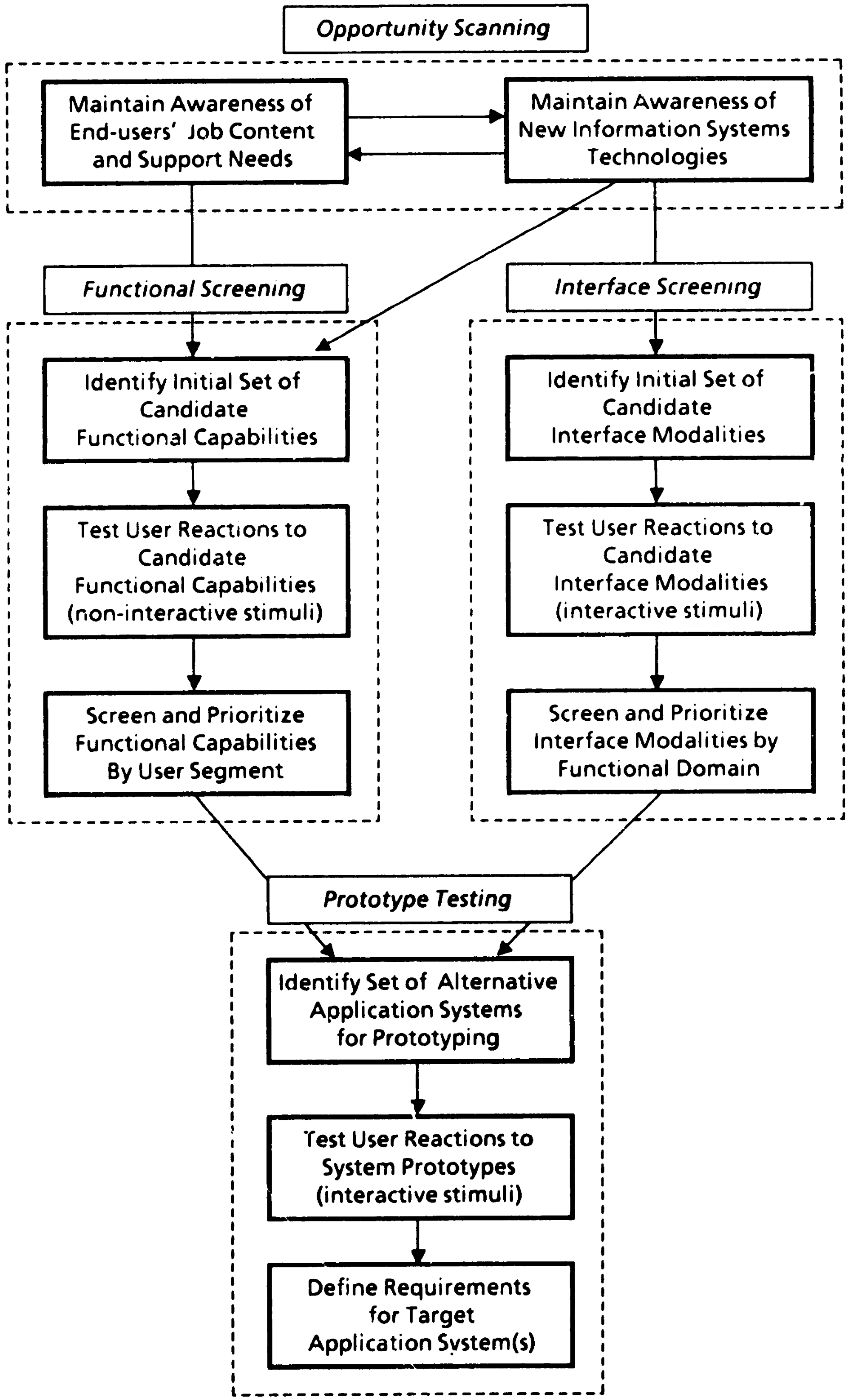
A Generic User Acceptance Testing Procedure

An important question raised by the reserch results reported above is how the technology acceptance model might potentially be applied in practical design settings. Figure 6.1 depicts a potential generic "user acceptance testing" procedure suggested by the present research. The objective of this procedure is to systematically select from the wide array of possible new support systems those which have the highest probability of meeting the needs of, and being accepted by, the intended users. The overall procedure consists of four distinct subprocedures each having different objectives: opportunity scanning, functional screening, interface screening, and prototype testing.

Opportunity Scanning

The objective of opportunity scanning is to create and maintain an up-to-date understanding of new and emerging information technologies, and an awareness of the applicability of those new technologies to the job content of potential end-users. In analyzing users' job content, emphasis must be place on the user and his or her job activities, and not on the technology per se. Thus, this element of opportunity scanning is similar in spirit to traditional systems analysis. A variety of tools specifically geared for analyzing and documenting the job content of potential end-users have beeen established over the past several years. The "critical success factors" approach (Rockart, 1979), for example, places the focus on the user by eliciting the factors that are most critical to the successful accomplishment of his or her business objectives. The "office automation methodology" (Sirbu, et al., 1983) provides a framework of descriptive concepts and an associated interviewing methodology to

Figure 6.1. Generic User Acceptance Testing Procedure



descriptively analyze the work activities of office workers as an input to specifying systems needs. Methods such as these may be readily adapted to the present user acceptance testing procedure for analyzing the job content of potential end users. The importance of understanding users job content on an on-going basis is underscored by the increasing recognition on behalf of senior managers of the value of linking information systems priorities to business priorities (e.g., Benjamin, et al., 1984; McFarlan & McKenney, 1983; Rockart, 1982).

In parallel with maintaining an awareness of users' job content, the Information Systems organization must keep up to date on new and emerging information technologies. The emergence of new information technologies can typically be predicted within a five year horizon, and often longer, with a fair degree of confidence by professionals with the appropriate expertise. Even after new technologies are available in the marketplace, there is generally a significant lag time before they experience widescale application. Organizations are increasingly recognizing the strategic advantages of being able to apply the latest technologies in an effective manner (e.g., Benjamin, et al., 1984; McFarlan, 1984). As these new technologies become available, organizations should be in a position to analyze their potential role in the company. Thus, there should be an active interplay between technology tracking and user job content analysis efforts, as depicted in Figure 6.1. Many organizations are establishing specific departments charged with the mission of identifying opportunities for effectively applying new information systems in the business (e.g., Benjamin, et al., 1984; McFarlan, 1984).

The combined awareness of users' job content and new information technologies permits the organization to proactively identify potential opportunities for efffective computer-based support of end-users.

Functional Screening

Functional screening aims to cull from among a large set of potential functional capabilities those which are perceived to be most useful to members of the potential user audience. A functional capability is the specification of the way in which some new or existing information technology can be applied to the job activities of a potential user group. The focus here is on measuring the perceived usefulness, rather than the perceived ease of use, of the candidate functional capabilities. An initial set of candidate functional capabilities is identified via the preceding opportunity scanning phase. The candidate functional capabilities are then presented to users via a non-interactive stimulus medium such as verbal description, slide presentation or videotape. The present research suggests that non-interactive media (in particular videotape) have promise for accurate assesment of perceived usefulness, although future research is needed to define the bounds of this accuracy and how it is affected by the specific characteristics of the non-interactive medium. Non-interactive media have numerous advantages relative to interactive media, as discussed earlier. They enable IS practitioners to evaluate functional capabilities which do not presently exist, and to administer studies across a wide user audience with a relatively low cost per subject. One characteristic that is likely to be important in the non-interactive medium is whether or not it presents the functional capability in terms of how it might be used by the target user in a work context, instead of simply presenting its technical aspects. Potential users may have difficulty judgeing the applicability to their jobs of very new technologies which they have never seen before and about which they have little knowledge. Presenting the new technical capability in terms of how it would be used in a work setting should make it easier for subjects to judge the benefits and drawbacks of the new capability relative to their job needs. It remains a research question as to how the accuracy of usefulness judgements is affected by the degree to which the technical capabilities being presented deviate from support tools with which the subjects are more familiar.

In addition to measuring the subjects' perceived usefulness of the candidate functional capabilities, the perceived importance of the tasks being supported to the subjects' jobs can also be measured. This would enable the IS practitioner to identify segments of the user population for which specific applications are especially useful, using cluster analysis (e.g., Wind, 1978), for example. The resulting data is used to prioritize the various functional capabilities and to screen out those viewed as low in perceived usefulness.

Interface Screening

Interface screening may be conducted independently of, and in parallel with, functional screening. New information technologies are rapidly creating opportunities for new and different modes of human-computer interaction such as: voice input and output, natural language input, touch screens, three-dimensional touch-sensitive tablets, animation, and eye-movement input. Interface screening seeks to identify the interface modalities that are easiest and most enjoyable for subjects to use in performing various computer-based tasks. The present research suggests that in order for subjects to form accurate ease of use perceptions, interface testing should be performed using direct hands-on interaction with test systems. There are many examples of tests of this kind in the human factors literature (see Chapter 3). Due to the relative expense of conducting such interface studies, resulting from the need for hands-on interaction with a test system interface, it is expected that a smaller number of subjects will be employed for interface screening than for functional screening. In addition, a single interface screening study may be used in conjunction with multiple functional screening studies. Results of the interface screening study are used to prioritize the various candidate interface modalities in terms of ease of use and enjoyability. In some cases, the prototype screening procedure may be profitably combined with the prototype testing procedure discussed below. This is particularly likely in cases where there are a small number of candidate interface modalities under consideration.

Prototype Testing

When the set of alternative functional capabilities and interface modalities have been prioritized and narrowed down to a few alternatives, the prototype testing phase is performed. First, a set of alternative sysems is defined by taking combinations of high priority functional capabilities and interface modalities. The alternative systems are prototyped, possibly taking advantage of various available rapid prototyping or system simulation tools (e.g., Beregi, 1984; Maurer, 1983). Next, users perform sample tasks using the prototype versions. Their motivational reactions to the system alternatives are measured, including constructs such as perceived usefulness, perceived ease of use, anticipated enjoyment of using, quality of system output, behavioral expectation and possibly task-specific perceptions and importance weights as well. These are discussed in greater detail in Chapter 5. The analysis of these user measurements permits one to identify the system or systems which are expected to be most acceptable to the target users. In addition the perceptual data give diagnostic information as to the reasons underlying the overall likelihood of acceptance, and may suggest other designs to be considered. As in other user testing paradigms (e.g., Gould & Lewis, 1983), designers may in some cases find it valuable to iteratively test additional configurations before settling on a final target system (or set of systems). The prototype alternative selected as the final target application system serves to define the requirements for the new system. System developers can then proceed to develop the final system based on these requirements according to existing development procedures such as the system development life cycle (e.g., Alavi, 1984).

Staffing and Organizing the User Acceptance Testing Function

It is important to consider the skills required of MIS professionals in order to perform user acceptance testing. Although further research and experience will be needed to gain a better understanding of these skill requirements, the various steps of the proposed user acceptance testing procedure suggest the kinds of expertise needed. Key skill areas needed to carry out the opportunity scanning procedure include: an ability to understand the fundamental business processes that form the job content of users, an ability to effectively interact with current or potential information systems users, and a fairly advanced level of expertise in information systems technology. Insofar as these skills are hard to find in single individuals, a multi-disciplinary team composition is indicated. For the screening and testing procedures, the additional research skills of experimental design, questionaire design, sampling, running subjects and analyzing data will be needed. Several issues surrounding how the user acceptance testing function should be organized will need to be better understood in order to put the proposed procedure into practice. Of particular concern are the division of responsibilities for the various procedures, and the ability of the user acceptance testing group to maintain an objective perspective, relatively insulated from the potentially dysfunctional influences of coalitions favoring a particular system configuration. It is likely that in many organizations there presently exist groups for whom the addition of user acceptance testing responsibilities would be a natural step. Already mentioned are the groups in many companies responsibile for technology scanning. In addition, "infocenters" and other forms of end-user support organizations often have responsibility for matching available tools to user needs. Finally, many vendor organizations have usability testing laboratories that perform a function similar in form to interface screening, although typically done at a much later point in the development process. These existing groups may form the basis for the user acceptance testing function.

Directions for Future Research

The purpose of this section is to briefly outline a series of topics for future reseearch on the theoretical and applied aspects of the technology acceptance model.

Subjective vs. objective ease of use. The technology acceptance model measures perceived ease of use, while numerous human factors approaches typically measure objective ease of use in the form of various laboratory performance metrics such as speed of task completion and error rate (see Chapter 3). A important question concerns the extent to which objective and subjective ease of use correlate. If they do not correlate in a given situation, which is the correct ease of use? are they different kinds of ease of use? The domains within which each type of ease of use measure are most appropriate would need to be defined. For example, objective ease of use may be more appropriate for non-discretionary systems such as order entry systems, while subjective ease of use may be more appropriate for discretionary systems for which the subjective reaction is a key determinant of the success of the system.

Subjective vs. objective usefulness. A similar analysis could be enlightening for usefulness. Are the expected gains in job performance predictive of the actual performance gains? Do users believe that they are more productive while available productivity data fails to bear this out? One barrier to research in this area is the formidable challange of operationalizing and defining organizational performance.

User adoption as a goal. The present model views the acceptance and use of a new system a behavior that is largely under the volitional control of the potential user. However, some users may regard the adoption of a new system as a goal, with some probability of failing to achieve the goal due to ability limitations. If so, an important research question is the extent to which expectancies and consequences of success and failure influence the motivation of users to attemtpt to adopt the target system. Recent theoretical extensions to the Fishbein model (Warshaw & Davis, 1985; in press; Warshaw, Sheppard & Hartwick, in press) are specifically geared to addressing behavioral goals. This recent theorizing and research provides an appropriate basis for addressing goal aspects of adoption.

Subjective Norm component. The subjective norm component of the Fshbein model was omitted from the versions of the technology acceptance model tested herein. However, this variable represents a potential source of increased explanatory power in predicting organizational adoption of systems. We have argues in Chapter 2 that subjective norms are not likely to be operative in the laboratory cpontext. Although TAM is intended to be employed in laboratory-based user acceptance tests, it is also intended to reflect organizationally-based acceptance processes outside the laboratory. Thus, it should be concerned with the role of social normative influences in the organizational environment. Thus it is important to direct future research attention to the role of subjective normative influences in user accepttance processes.

Recall that the objective of the model is to explain the causal mechanisms linking the design characteristics of systems to actual usage behavior. Thus, some theoretical concern centers around the possible role of subjective norm as an alternative mechanism by which differences in system features may affect usage. It is quite plausible that the characteristics of a system may affect a referent's opinion regarding whether or not a potential user should or should not use that system. If such an effect on subjective norm directly influences intention or behavior, then we should view subjective norm as a mediating construct apart from attitude. Conversely, if social norm influences behavior only indirectly though its effect on attitude, then subjective norm does not function as an independent mediator, it's effects on behavior being mediated by attitude.

Although indirect effect of subjective norm on intention through attitude is not explicitly dealt with by the Fishbein model, recent research has suggested that such effects may have a major role. The conceptual foundation for this work is Kelman's (1961) theoretical distinction between three different processes by which societal influences may effect behavioral: identification, internalization, and compliance. Identification and internalization effects of social influence are theorized to operate through the individual's attitudinal structure (Warshaw, 1980). Compliance refers to situations where the individual performs a behavior which is inconsistent with his or her own attitude in order to gain rewards and avoid punishments from salient referents, and is associated with "felt pressure" (Warshaw, 1980). Ajzen and Fishbein (1980, p. 262) suggest measuring subjective norm using scale wording such as "Most people who are important to me think (I should/should not) perform behavior x." As Warshaw (1980) points out, this may not reflect the true influence of referents on one's intent, since: "the subject may want to do what Referent X thinks he/she should do, not because of X's influence, but because the act is consistent with the subject's own A_b." In fact, this is the interpretation often attributed to the high correlations typically observed between attitudes and norms, which often exceed the correlations between either of these constructs and intention (Warshaw, 1980). Thus, it appears that the standard subjective norm scale picks up the coincidental equivalence of referent expectations and subject attitude, as well as the internalization, identification, and compliance elements of true social influence processes. The implications for the present research are that internalization and identification processes of social influence may represent mechanisms linking system features to an individual's attitudinal belief structure, whereas compliance may represent an alternative normative mediator apart from attitude.

Future research on the role of subjective normative influences relative tot he existing technology acceptance models may yield a more complete understanding of the dynamics of user acceptance processes.

Conclusion

Managerial Importance of Thesis

The research presented above is important to the concerns of managers from three perspectives: (1) from the perspective of the manager as a potential user of a new information system, (2) from the perspective of the manager of the design team or organization responsible for developing new end-user information systems, (3) from the perspective of the manager of the user organization.

Manager as User. Managers represent a major component of the overall population of potential end-users. The job content of managers has a large emphasis on information processing, communication and decision-making activities (e.g., Mintzberg, 1973). These are all activities for which computer-based support can play a useful role. A great deal of information is needed to plan and coordinate the activities of the resources and people under the manager's responsibility. The manager is frequently faced with high levels of uncertainty for which access to the proper information and to the appropriate tools for analyzing, summarizing, interpreting and displaying the information is of value. The idea that managers could benefit from the direct use of computer-based information systems has been around for some time. Due to the importance of what managers do, and the high salaries they are paid, their support through information systems has been a high priority. Unfortunately, past progress in the development of computer support systems has been slow and painful. Reasons for this include the high complexity of managerial work, limitations in the technical tools available, and lack of understanding of how to design effective support systems for managers (Keen & Scott Morton, 1978). More recently, a fairly dramatic upsurge in the effective design and implementation of managerial support systems has taken place (Rockart & Scott Morton, 1984; Rockart & Treacy, 1982; Keen & Woodman, 1984). The advances in personal computing hardware and software have been a major influence in this growth. In addition, the increasingly sophisticated techniques being used to understand the nature of managerial work and the process of designing effective systems to support have been factors. It is expected that the user acceptance testing methods presented in the present resdearch will contribute to our ability to design increasingly useful systems for managerial end-users.

Management of The Development Organization. Managers responsible for directing the activities of the design or development organization attempting to create successul new systems should also benefit from the above research. The technology acceptance model enables the development manager to better understand the key determinants of user acceptance of new systems, and to understand how various key decisions of the design team may effect the success of the new systems they generate. In this context the model serves as a framework for thinking through and establishing the varisous requirements and design criterai for a given new system. The proposed user acceptance testing process enables the manager to evaluate proposed new technologies as they progress, and to guide the efforts of the design team toward high priority design configurations. Senior managers of development organizations are calling for developers to pay greater attention to designing systems which are easier to use and more useful, and to employ techniques for testing and refining systems during the development process (Branscomb & Thomas, 1984; Gould & Lewis, 1983). The present research makes a contribution to the set of techniques available for assuring highly acceptable systems. The proposed user acceptance testing process is likely to take a great deal of the guess-work out of managing development, and to reduce the risk of costly implementation or market failure of new systems.

Management of the User Organization. General managers are finding themselves playing an increasing role in decisions regarding the computer support of their subordinates. Two key questions the manager must face are (1) how do I select a system that fits the needs of the end-users? and (2) is the selected new system justifiable in terms of the benefits expected? The present research makes a contribution toward both of these questions. In many cases there is substantial uncertainty on the part of the users as to what systems would best fit their needs. A major existing approach to information systems development is for systems analysts to interview the users about their work activities and support requirements, and then use that information to define the requirements for a new system. The systems analysis process is similar to the "opportunity scanning" process of the suggested user acceptance testing procedure (Figure 6.1). A key difference is that, whereas today the systems analyst defines requirements directly from user interview data, the proposed user acceptance testing procedure includes several additional user testing steps between user interviews ("opportunity scanning") and requirements definition (the final step). These testing steps are aimed at ensuring the fit between the resulting system and the users needs, and hence should provide advantages relative to existing systems analysis approaches. Regarding system justification, managers are increasingly acknowledging the difficulty in objectively defining system benefits, adopting the alternative view that intangible "soft" benefits should be understood, measured, and predicted in order to establish a business case for a new system. The present research adds to the set of existing tools applicable to the task of measuring subjective (perceived) user benefits. Thus the technology acceptance model and associated user acceptance testing procedure offer advantages to general managers for managing the use of end-user systems by their subordinates.

Contribution Toward a Motivational Model of User

The user acceptance testing procedures discussed in this research represent a potentially valuable addition to the set of existing tools for aiding the system design and development process for end-user systems. A valid motivational model of the user that reflects the impact of design choices on user motivation is a key element in the success of user acceptance testing procedures. The technology acceptance model proposed and tested above represents a significant contribution toward establishing a valid motivational model of the user. Thus the present research has taken the first several steps toward the establishing a valid motivational model of the user. Moreover, this research has created a research foundation upon which investigators may base future research directed toward further progress in the understanding of user acceptance.

References

Ajzen, I. & Fishbein, M. (1970). The prediction of behavior from atttudinal and normative variables. Journal of Experimental Social Psychology, 6, 466-487.

Ajzen, I. & Fishbein, M. (1972). Attitudes and normative beliefs as factors influencing behavioral intentions. Journal of Personality and Social Psychology, 21, 1-9.

Ajzen, I. & Fishbein, M. (1973). Attitudinal and normative variables as predictors of specific behaviors. Journal of Personality and Social Psychology, 27, 41-57.

Ajzen, I. & Fishbein, M. (1977). Attitude-behavior relations: A theoretical analysis and review of empirical research. Psychological Bulletin, 84, 888-918.

Ajzen, I. & Fishbein, M. (1980). Understanding attitudes and predicting social behavior. Englewood Cliffs, NJ: Prentice-Hall.

Alavi, M. (1984). An assessment of the prototyping approach to information systems development. Communications of the ACM, 27, 556-563.

Alavi, M. & Henderson, J.C. (1981). An evolutionary strategy for implementing a decision support system. Management Science, 27, 1309-1323.

Anderson, N.H. (1970). Functional measurement and psycho-physical judgement. Psychological Review, 77, 153-170.

Asch, S.E. (1946). Forming impressions of personality. Journal of Abnormal and Social Psychology, 41, 258-290.

Atkinson, J.W. (1957). Motivational determinants of risk-taking behavior. Psychological Review, 64, 359-372.

Bagozzi, R.P. (1981). Attitudes, intentions and behaviors: A test of some key hypotheses. Journal of Personality and Social Psychology, 41, 607-627.

Bagozzi, R.P. (1982). A field investigation of causal relations among cognitions, affect, intentions and behavior. Journal of Marketing Research, 19, 562-584.

Bagozzi, R.P. (1983). A holistic methodology for modeling consumer response to innovation. Operations Research, 31, 128-176.

Bagozzi, R.P. (1984). Expectancy-value attitude models: An analysis of critical measurement issues. International Journal of Research in Marketing, 1, 295-310.

Bailey, J.E. & Pearson, S.W. (1983). Development of a tool for measuring and analyzing computer user satisfaction. Management Science, 29, 530-545.

Barber, R.E. & Lucas, H.C. (1983). System response time, operator productivity, and job satisfaction. Communications of the ACM, 26, 9, 972-986.

Barrett, G.V., Thornton, C.L., & Cabe, P.A. (1968). Human factors evaluation of a computer based storage and retrieval system. Human Factors, 10, 4 431-436.

Bass, F.M. & Wilkie, W.L. (1973). A comparative analysis of attitudinal predictions of brand preference. Journal of Marketing Research, 10, 262-269.

Benbasat, I. & Schroeder, R.G. (1977). An experimental investigation of some MIS design variables. MIS Quarterly, 1, 37-49.

Benjamin, R.I. (1982). Information technology in the 1990's: A long range planning scenario. MIS Quarterly, 6, 11-31.

Benjamin, R.I., Rockart, J.F., Scott Morton, M.S., & Wyman, J. (1984). Information Technology: A Strategic Opportunity. Sloan Management Review, 25, 3, 3-10.

Bentler, P.M. & Speckart, G. (1979). Models of attitude-behavior relations. Psychological Review, 86, 452-464.

Beregi, W.E. (1984). Architecture prototyping in the software engineering environment. IBM Systems Journal, 23, 4-18.

Bewley, W.L., Roberts, T.L., Schoit, D., & Verplank, W.L. (1983). Human factors testing in the design of Xerox's 8010 "star" office workstation. CHI '83 Human Factors in Computing Systems (Boston, December 12-15, 1983), ACM, New York, 72-77.

Bohrnstedt, G.W. (1970). Reliability and validity assessment in attitude measurement. In G.F. Summers (Ed.), Attitude measurement. Chicago: Rand-McNally, Ch. 3, 80-99.

Branscomb, L.M. & Thomas, J.C. (1984). Ease of use: A system design challange. IBM Systems Journal, 23, 224-235.

Brinberg, D. (1979). An examination of the determinants of intention and behavior: A comparison of two models. Journal of Applied Social Psychology, 9, 560-575.

Brinberg, D. & Durand, J. (1983). Eating at fast-food restaurants: A comparison of two models. Journal of Applied Social Psychology, 9, 560-575.

Brosey, M. & Shneiderman, B. (1978). Two experimental comparisons of relational and hierarchical database models. International Journal of Man-machine Studies, 10, 625-637.

Brunswik, E. (1955). Representative design and probability theory in a functional psychology. Psychological Review, 62, 193-217.

Butler, T.W. (1983). Computer response time and user performance. CHI '83 Human Factors in Computing Systems (Boston, December 12-15, 1983), ACM, New York, 58-62.

Calder, B. J. (1977). Focus groups and the nature of qualitative marketing research. Journal of Marketing Research, 14, 353-364.

Campbell, D.T. & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 56, 81-105.

Campbell, D.T., Siegman, C.R. & Rees, M.B. (1967). Direction-of-wording effects in the relationships between scales. Psychological Bulletin, 68, 293-303.

Card, S.K., English, W.K., & Burr, B.J. (1978). Evaluation of mouse, rate-controlled isometric joystick, step keys, and text keys for text selection on a CRT. Ergonomics, 21, 601-613.

Card, S.K., Moran, T.P., & Newell, A. (1980). The keystroke-level model for user performance time with interactive systems. Communications of the ACM, 23, 396-410.

Chervany, N.L. & Dickson, G.W. (1974). An experimental evaluation of information overload in a production environment. Management Science 18, 113-133.

Churchill, G.A. (1979). A paradigm for developing better measures of marketing constructs. Journal of Marketing Research, 16, 64-73.

Cook, D. (1967). The impact of the Hawthorne effect in experimental designs in educational research. Washington, D.C.: U.S. Government Printing Office.

Cook, T.D. & Campbell, D.T. (1979). Quasi-experimentation: Design and analysis issues for field settings. Boston: Houghton Mifflin.

Cronbach, L.J. (1951). Coefficient aplha and the internal structure of tests. Psychometrika, 16, 297-334.

Cronbach, L.J. & Meehl, P.E. (1955). Construct validity in psychological tests. Psychological Bulletin, 52, 281-302.

Davidson, A.R. & Jaccard, J.J. (1979). Variables that moderate the attitude-behavior relation: Results of a longitudinal survey. Journal of Personality and Social Psychology, 37, 1364-1376.

Davidson, A., & Morrison, D. (1983). Predicting contraceptive behavior from attitudes: A comparison of within- versus across-subjects procedures. Journal of Personality and Social Psychology, 45, 997-1009.

Davis, G.B. (1982). Strategies for information requirements determination. IBM Systems Journal, 21, 4-30.

Dawes, R.M. & Corrigan, B. (1974). Linear models in decision making. Psychological Bulletin, 81, 95-106.

Deci, E.L. (1975). Intrinsic Motivation. New York: Plenum Press.

DeSanctis, G. (1983). Expectancy theory as an explanation of voluntary use of a decision support system. Psychological Reports, 52, 247-260.

Dickson, G.W., Senn, J.A., & Chervany, N.L. (1977). Research in management information systems: The Minnesota experiments. Management Science, 23, 913-923.

Doob, L.W. (1947). The behavior of attitudes. Psychological Review, 54, 135-156.

Dulaney, D.E. (1961). Hypotheses and habits in verbal "operant conditioning." Journal of Abnormal and Social Psychology, 63, 251-263.

Duncan, O.D. (1975). Introduction to structural equation models. New York: Academic Press.

Dzida, W., Herda, S. & Itzfeldt, W.D. (1978). User-perceived quality of interactive systems. IEEE Transactions on Software Engineering, SE-4, 270-276.

Edwards, W. (1954). The theory of decision making. Psychological Bulletin, 51, 380-417.

Einhorn, H.J., & Hogarth, R.M. (1975). Unit weighting schemes for decision making. Organizational Behavior and Human Performance, 13, 171-192.

Einhorn, H.J. & Hogarth, R.M. (1981). Behavioral decision theory: Processes of judgement and choice. Annual Review of Psychology, 32, 53-88.

Einhorn, H.J., Kleinmuntz, D.N. & Kleinmuntz, B., (1979). Linear regression and process-tracing model of judgement. Psychological Review, 86, 465-485.

Embley, D.W. & Nagy, G. (1981). Behavioral aspects of text editors. Computing Surveys, 13, 33-70.

Fazio, R.H. & Zanna, M.P. (1978). Attitudinal qualities relating to the strength of the attitude-behavior relationship. Journal of Experimental Social Psychology, 14, 398-408.

Fazio, R.H. & Zanna, M.P. (1981). Direct experience and attitude-behavior consistency. Advances in Experimental Social Psychology, 14, 161-202.

Festinger, L.A. (1957). A theory of cognitive dissonance. Evanston, IL: Row, Peterson.

Fishbein, M. (1963). An investigation of the relationships between beliefs about an object and the attitude toward that object. Human Relations, 16, 233-240.

Fishbein, M. (1964). The relationship of the behavioral differential to other attitude instruments. American Psychologist, 19, 540.

Fishbein, M. (1966). Sexual behavior and propositional control. Paper read at the Psychonomic Society meeting.

Fishbein, M. (1967). Attitude and the prediction of behavior. In Fishbein, M. (Ed.) Readings in attitude theory and measurement. New York: Wiley.

Fishbein, M. & Ajzen, I. (1974). Attitudes toward objects as predictors of single and multiple behavioral criteria. Psychological Review, 81, 59-74.

Fishbein, M. & Ajzen, I. (1975). Belief, attitude, intention and behavior: An introduction to theory and research. Reading, MA: Addison-Wesley.

Fishbein, M. & Raven, B.H. (1962). The AB scales: An operational definition of belief and attitude. Human Relations, 15, 35-44.

Fudge, W.K. & Lodish, L.M. (1977). Evaluation of the effectiveness of a model based salesman's planning system by field experimentation. Interfaces, 8, 97-106.

Fuerst, W.L. & Cheney, P.H. (1982). Factors affecting the perceived utilization of computer-based decision support systems in the oil industry. Decision Sciences, 13, 554-569.

Gallagher, C.A. (1974). Perceptions of the value of a management information system. Academy of Management Journal, 17, 46-55.

Ghani, J.A. (1981). The effects of information representation and modification on decision performance. Unpublished doctoral dissertation, University of Pennsylvania.

Ghani, J.A. & Lusk, E.J. (1982). The impact of a change in information representation and a change in the amount of information on decision performance. Human Systems Management, 3, 270-278.

Ginzberg, M.J. (1981). Early diagnosis of MIS implementation failure: Promising results and unanswered questions. Management Science, 27, 459-478.

Good, M. (1982). An ease of use evaluation of an integrated document processing system. CHI '83 Human Factors in Computing Systems (Boston, December 12-15, 1983), ACM, New York, 142-147.

Gould, J.D., Conti, J., & Hovanyecz, T. (1983). Composing letters with a simulated listening typewriter. Communications of the ACM, 26, 295-308.

Gould, J.D. & Lewis, C. (1983). Designing for usability- key principles and what designers think. CHI '83 Human Factors in Computing Systems (Boston, December 12-15, 1983), ACM, New York, 50-53.

Greenblatt, D. & Waxman, J. (1978). A study of three database query languages. In B. Shneiderman, (Ed.) Databases: Improving usability and responsiveness. New York: Academis Press, 77-79.

Hammond, K.R., & Summers, D.A. (1972). Cognitive control. Psychological Review, 79, 58-67.

Hauser, J.R. & Simmie, P. (1981). Profit maximizing perceptual positions: An integrated theory for the selection of product features and price. Management Science, 27, 33-56.

Hauser, J.R. & Urban, G.L. (1977). A normative methodology for modeling consumer response to innovation. Operations Research, 25, 579-619.

Heider, G. (1946). Attitudes and cognitive organization. Journal of Psychology, 21, 107-112.

Heider, F. (1958). The psychology of interpersonal relations. New York: Wiley.

Henderson, J.H. & Nutt, P.C. (1980). The influence of decision style on decision making behavior. Management Science, 26, 371-386.

Holbrook, M.B. (1981). Integrating compositional and decompositional analyses to represent the intervening role of perceptions in evaluative judgements. Journal of Marketing Research, 18, 13-28.

Hom, P.W., Katerberg, R., & Hulin, C.L. (1979). Comparative examination of three approaches to the prediction of turnover. Journal of Applied Psychology, 64, 280-290.

Huber, G.P. (1983). Cognitive style as a basis for MIS and DSS design: Much ado about nothing? Management Science, 29, 567-582.

Igersheim, R.H. (1976). Managerial response to an information system. Proceedings of the National Computer Conference, Arlington, VA: AFIPS Press, 877-882.

Ives, B., Olson, M.H., & Baroudi, J.J. (1983). The measurement of user information satisfaction. Communications of the ACM, 26, 785-793.

Jaccard, J., & Davidson, A. (1972). Toward an understanding of family planning behaviors: An initial investigation. Journal of Applied Social Psychology, 2, 228-235.

Jaccard, J., & Davidson, A. (1975). A comparison of two models of social behavior: Results of a survey sample. Sociometry, 38, 497-517.

Jaccard, J. & Sheng, D. (1985). A comparison of six methods for assessing the importance of perceived consequences in behavioral decisions: Applications from attitude research. Journal of Experimental Social Psychology, 20, 1-28.

Jaccard, J., Weber, J., & Lundmark, J. (1975). A multitrait-multimethod analysis of four attitude assessment procedures. Journal of Experimental Social Psychology, 11, 149-154.

Janis, I.L. & Mann, L. (1977). Decision making: A psychological analysis of conflict, choice, and commitment. New York: Free Press.

Johanson, R. & Baker, E. (1984). User needs workshops: A new approach to anticipating user needs for advanced office systems. Office: Technology and People, 2, 103-119.

Johnson, E.J., & Meyer, R.J. (1984). Compensatory choice models of non-compensatory processes: The effect of varying context. Journal of Consumer Research, 11, 528-541.

Johnson, E.J. & Payne, J.W. (1985). Effort and accuracy in choice. Management Science, 31, 395-414.

Johnson, S.C. (1967). Hierarchial clustering schemes. Psychometrika, 32, 241-254.

Jones, E.E. & Davis, K.E. (1965). From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), Advances in Experimental Social Psychology, Vol. 2. New York: Academis Press, 219-266.

Kaiser, K. & Srinivasan, A. (1982). User-analyst differences: An empirical investigation of attitudes related to system development. Academy of Management Journal, 25, 630-646.

Keen, P.G.W. (1980). MIS Research: Reference disciplines and a cumulative tradition. Proceedings of the First International Conference on Information Systems, E.R. McLean (Ed.), Dec 8-10, Philadelphia, PA.

Keen, P.G.W. (1981). Value analysis: Justifying decision support systems. MIS Quarterly, 5, 1-15.

Keen, P.G.W. & Woodman, L. (1984). What To Do With All Those Micros. Harvard Business Review, September-October, 142-150.

Keen, P.G.W. & Scott Morton, M.S. (1978). Decision support systems: An organization perspective. Reading, MA: Addison-Wesley.

Kelley, H.H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), Nebraska Symposium on Motivation. Lincoln: University of Nebraska Press, 192-238.

Kelley, H.H. (1973). The process of causal attribution. American Psychologist, 28, 107-128.

Kelly, G.A. (1955). The psychology of personal constructs, vol. 1, New York: Norton.

Kelman, H.C. (1961). Processes of opinion change. Public Opinion Quarterly, 25, 57-78.

King, W.R. & Epstein, B.J. (1983). Assessing information system value: An experimental study. Decision Sciences, 14, 34-45.

King, W.R. & Rodriguez, J.I. (1981). Participative design of strategic decision support systems: An empirical assessment. Management Science, 27, 717-726.

Land, K. C. (1973). Identification, parameter estimation, and hypothesis testing in recursive sociological models. In Goldberger, A.S. & Duncan, O.D. (Eds.) Structural equation models in the social sciences. New York: Seminar Press, 19-49.

Larcker, D.F. & Lessig, V.P. (1980). Perceived usefulness of information: A psychometric examination. Decision Sciences, 11, 121-134.

Ledgard, H., Whiteside, J.A., Singer, A., & Seymour, W. (1980). The natural language of interactive systems. Communications of the ACM, 23, 556-563.

Lochovsky, F.H. & Tsichritzis, D.C. (1977). User performance considerations in DBMS selection. Proceedings of the ACM SIG-MOD, 128-134.

Lucas, H.C. (1975). Performance and the use of an information system. Management Science, 21, 908-919.

Lucas, H.C. (1978). Unsuccessful implementation: The case of a computer-based order entry system. Decision Sciences, 9, 68-79.

Lucas, H.C. (1981). An experimental investigation of the use of computer-based graphics in decision making. Management Science,Science, 27, 757-768.

Lucas, H.C. & Neilson, N.R. (1980). The impact of the mode of information presentation on learning and performance. Management Science, 26, 982-993.

Lusk, E.J. & Kersnick, M. (1979). The effect of cognitive style and report format on task performance: The MIS design consequences. Management Science, 25, 787-798.

Mandler, G. (1967). Verbal learning. In Newcomb, T. M. (Ed.) New Directions in Psychology, 3, New York: Holt, 1-50.

McArdle, J.B. (1972). Positive and negative communications and subsequent attitude and behavior change in alcoholics. Unpublished doctoral dissertation, University of Illinois.

McFarlan, F.W. (1984). Information Technology Changes the Way You Compete. Harvard Business Review, May-June, 98-103.

McFarlan, F.W. & McKenney, J.L. (1983). The Information Archipelago-Governing the New World. Harvard Business Review, July-August, 91-99.

McKennell, A.C. (1974). Surveying attitude structures: A discussion of principles and procedures. Quantity and Quality, 7, 203-293.

Magers, C.S. (1983). An experimental evaluation of on-line help for non-programmers. CHI '83 Human Factors in Computing Systems (Boston, December 12-15, 1983), ACM, New York, 277-281.

Maish, A.M. (1979). A user's behavior toward his MIS. MIS Quarterly, 3, 39-52.

Maguire, M. (1982). An evaluation of published recommendations on the design of man-computer dialogues. International Journal of Man-machine Studies, 16, 237-261.

Malone, T.W. (1981). Toward a theory of intrinsically motivating instruction. Cognitive Science, 4, 333-369.

Mandler, G. Verbal learning. In T.M. Newcomb (Ed.), New Directions in Psychology, Vol. 3, New York, Holt, 1-50.

Manstead, A.S.R., Proffitt, C. & Smart, J.L. (1983). Predicting and understanding mothers' infant-feeding intentions and behavior: Testing the theory of reasoned action. Journal of Personality and Social Psychology, 44, 657-671.

Markus, M.L. (1983). Power, politics, and MIS implementation. Communications of the ACM, 26, 430-444.

Maurer, M.E. (1983). Full-screen testing of interactive applications. IBM Systems Journal, 22, 246-261.

Miller, G.A. (1956). The magical number seven; plus or minus two: Some limits on our capacity for processing information. Psychological Review, 63, 81-97.

Miller, L.H. (1977). A study in man-machine interaction. National Computer Conference, 409-421.

Mintzberg, H. (1973). The nature of managerial work. New York: Harper & Row.

Newell, A. & Simon, H.A. (1972). Human problem solving. Englewood Cliffs, NJ: Prentive-Hall.

Nisbett, R.E., & Wilson, T.D. (1977). Telling more than we can know: Verbal reports on mental processes. Psychological Review, 84, 231-259.

Norman, D.A. (1983). Design principles for human-computer interfaces. CHI '83 Human Factors in Computing Systems (Boston, December 12-15, 1983), ACM, New York, 1-10.

Nunnally, J. C. (1978). Psychometric Theory. New York: McGraw-Hill.

Osgood, C.E., Suci, G.J., & Tannenbaum, P.H. (1957). The measurement of meaning. Urbana, University of Illinois Press.

Ostrom, T.M. (1969). The relationship between the affective, behavioral, and cognitive components of attitude. Journal of Experimental Social Psychology, 5, 12-30.

Pindyck, R.S., & Rubinfeld, D.L. (1981). Econometric models and economic forecasts. New York: McGraw-Hill.

Poller, M.F. & Garter, S.K. (1983). A comparative study of moded and modeless text editing by experienced editor users. CHI '83 Human Factors in Computing Systems (Boston, December 12-15, 1983), ACM, New York, 166-170.

Poppel, H.L. (1982). Who needs the office of the future? Harvard Business Review, 60, 146-155.

Powers, R.F. & Dickson, G.W. (1973). MIS project management: Myths, opinions, and reality. California Management Review, 15, 147-156.

Price, L.A. & Cordova, C.A. (1983). Use of mouse buttons. CHI '83 Human Factors in Computing Systems (Boston, December 12-15, 1983), ACM, New York, 262-266.

Reisner, P. (1981). Human factors studies of database query languages: A survey and assessment. Computing Surveys, 13, 13-31.

Reisner, P. Boyce, R.F. & Chamberlin, D.D. (1975). Human Factors evaluation of two database query languages: SQUARE and SEQUEL. Proceedings of the National Computer Conference. Montvale, NJ: AFPIS Press, 447-452.

Remus, W. (1984). An empirical investigation of the impact of graphical and tabular data presentations on decision making. Management Science, 30, 533-542.

Rice, R.E. (1980). The impact of computer mediated organizational and interpersonal communication. In Williams, M. (Ed.) Annual Review of Information Science and Technology, 15, White Plains, NY: Knowledge Industries, 221-249.

Roberts, T.L. & Moran, T.P. (1983). The evaluation of text editors: Methodology and empirical results. Communications of the ACM, 26, 265-283.

Robey, D. (1979). User attitudes and management information system use. Academy of Management Journal, 22, 527-538.

Robey, D. & Zeller, R.L. (1978). Factors affecting the success and failure of an information system for product quality. Interfaces, 8, 70-75.

Robinson, J.P. & Shaver, P.R. (1969). Measures of social psychological attitudes. Ann Arbor, MI: Institute for Social Research.

Rockart, J.F. (1979). Chief executives define their own data needs. Harvard Business Review, 57, 81-91.

Rockart, J.F. (1982). The changing role of the information systems executive: A critical success factors perspective. Sloan Management Review. 24, 3-13.

Rockart, J.F. & Flannery, L.S. (1983). The management of end user computing. Communications of the ACM, 26, 776-784.

Rockart, J.F. & Scott Morton, M.S. (1984). Implications of changes in information technology for corporate strategy. Interfaces, 14, 84-95.

Rockart, J.F. & Treacy, M.E. (1982). The CEO goes online. Harvard Business Review, January/February 1982, 82-88.

Rosenberg, M.J. (1956). Cognitive structure and attitudinal affect. Journal of Abnormal and Social Psychology, 53, 367-372.

Rosenberg, M.J. (1960). An analysis of affective-cognitive consistency. In C.I. Hovland & M. J. Rosenberg (Eds.) Attitude organization and change. New Haven: Yale University Press, 1960, 15-64.

Rosenberg, M.J. (1969). The conditions and consequences of evaluation apprehension. In Rosenthal, R. & Rosnow, R.L. (Eds.) Artifact in behavioral research. New York: Academic Press.

Rosenthal, R. (1972). On the social psychology of the self-fulfilling prophecy: Further evidence for Pygmalion effects and their mediating mechanisms. Unpublished manuscript, Harvard University.

Rotter, J.B. (1954). Social learning and clinical psychology. Englewood Cliffs, NJ: Prentice-Hall.

Ryan, M.J. (1982). Behavioral intention formation: A structural equation analysis of attitudinal and social influence interdependency. Journal of Consumer Research, 9, 263-278.

Ryan, M.J., & Bonfield, E.H. (1975). The Fishbein extended model and consumer behavior. Journal of Consumer Research, 2, 118-136.

Saltzer, E. (1981). Cognitive moderators of the relationship between behavioral intentions and behavior. Journal of Personality and Social Psychology, 41, 260-271.

Schewe, C.D. (1976). The management information system user: An exploratory behavioral analysis. Academy of Management Journal, 19, 577-590.

Schmidt, F.L. (1973). Implications of a measurement problem for expectancy theory research. Organizational Behavior and Human Performance, 10, 243-251.

Schultz, R.L. & Slevin, D.P. (1975). In Schultz, R.L. & Slevin, D.P. (Eds.) Implementing operations research/ management science. New York: American Elsevier, 153-182.

Shaw, M.E. & Wright, J.M. (1967). Scales for the measurement of attitudes. New York: McGraw-Hill.

Sherif, M. & Sherif, C.W. (1967). The own categories procedure in attitude research. In M. Fishbein (Ed.) Readings in attitude theory and measurement. New York: Wiley, p. 190-198.

Shneiderman, B., Mayer, R., McKay, D. & Heller, P. (1977). Experimental investigations of the utility of detailed flowcharts in programming. Communications of the ACM, 20, 373-381.

Shocker, A.D. & Srinivasan, V. (1979). Multiattribute approaches for product concept evaluation and generation: A critical review. Journal of Marketing Research, 16, 159-180.

Silk, A.J. (1969). Preference and Perception measures in new product development: An exposition and review. Industrial Management Review, Fall 1969, 21-37.

Silk, A.J. (1971). Response set and measurement of self-designated opinion leadership. Public Opinion Quarterly, 35, 383-397.

Silk, A.J. & Urban, G.L. (1978). Pre-test-market evaluation of new packaged goods: A model and measurement methodology. Journal of Marketing Research, 15, 171-191.

Simon, H.A. (1954). Spurious correlation: A causal interpretation. Journal of the American Statistical Association, 49, 467-479.

Sirbu, M.A., Schoichet, S.R., Kunin, J.S., Hammer, M.M., Sutherland, J.B. & Zarmer, C.L. (1983). Office analysis: Methodology and case studies. Cambridge, MA: Massachusetts Institute of Technology.

Slovic, P. & Lichtenstein, S. (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgement. Organizational Behavior and Human Performance, 6, 649-744.

Staats, A.W. & Staats, C.K. (1958). Effect of number of trials on the language conditioning of meaning. Journal of General Psychology, 61, 211-223.

Swanson, E.B. (1974). Management information systems: Appreciation and Involvement. Management Science, 21, 178-188.

Swanson, E.B. (1982). Measuring user attitudes in MIS research: A review. OMEGA, 10, 157-165.

Thomas, J.C. & Gould, J.D. (1975). A psychological study of Query by Example. Proceedings of the National Computer Conference. Montvale, NJ: AFIPS Press, 439-445.

Tolman, E.C. (1932). Purposive behavior in animals and men. New York: Appleton-Century-Crofts.

Triandis, H.C. (1977). Interpersonal behavior. Monterey, CA: Brooks/Cole.

Tybout, A.M. & Hauser, J.R. (1981). A marketing audit using a conceptual model of consumer behavior: Application and evaluation. Journal of Marketing, 45, 82-101.

Vroom, V.H. (1964). Work and motivation. New York: Wiley.

Wainer, H. (1976). Estimating coefficients in linear models: It don't make no nevermind. Psychological Bulletin, 83, 213-217.

Warshaw, P.R. (1980a). Predicting purchase and other behaviors from general and contextually specific intentions. Journal of Marketing Research, 17, 26-33.

Warshaw, P.R. (1980b). A new model for predicting behavior intentions: An alternative to Fishbein. Journal of Marketing Research, 17, 153-172.

Warshaw, P.R. & Davis, F.D. (1984). Self-understanding and the accuracy of behavioral expectations. Personality and Social Psychology Bulletin, 10, 111-118.

Warshaw, P.R. & Davis, F.D. (1985). Disentangling behavioral intention and behavior expectation. Journal of Experimental Social Psychology, 21, 213-228.

Warshaw, P.R. & Davis, F.D. (in press). The accuracy of behavioral intention versus behavioral expectation for predicting behavioral goals. Journal of Psychology.

Warshaw, P.R., Sheppard, B.H., & Hartwick, J. (in press). The intention and self-prediction of goals and behavior. In R. Bagozzi (Ed.), Advances in marketing communication. Greenwich, CT: JAI Press.

Weber, S.J. & Cook, D.T. (1972). Subject effects in laboratory research: An examination of subject roles, demand characteristics, and valid inference. Psychological Bulletin, 77, 273-295.

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. Psychological Review, 92, 548-573.

Weizer, N. & Jackson, L. (1983). The U.S. office automation market, 1983-1988: The evolution toward integrated information processing. Cambridge, MA: Arthur D. Little, Inc.

Welty, C., & Stemple, D.W. (1981). Human factors comparison of a procedural and a nonprocedural query language. ACM Transactions on Database Systems.

Wicker, A.W. (1969). Attitudes vs. actions: The relationship of verbal and overt behavioral responses to attitude objects. Journal of Social Issues, 25, 41-78.

Wilkie, W.L. & Pessemier (1973). Issues in marketing's use of multi-attribute attitude models. Journal of Marketing Research, 10, 428-441.

Wind, Y. (1978). Issues and advances in segmentation research. Journal of Marketing Research, 15, 317-337.

*Woodworth, R.S. & Schlosberg, H. (1954). Experimental psychology. New York: Holt.

Wyer, R.S. & Goldberg, L. (1970). A probabilistic analysis of relationships among beliefs and attitudes. Psychological Review, 77, 100-120.

Zmud, R.W. (1978). An empirical investigation of the dimensionality of the concept of information. Decision Sciences, 9, 187-195.

Zmud, R.W. (1979). Individual differences and MIS success: A review of the empirical literature. Management Science, 25, 966-979.

Zmud, R.W., Blocher, E. & Mofffie, R.P. (1983). The impact of color graphic report formats on decision performance and learning. Proceedings of the Fourth International Conference on Information Systems (December 15-17, 1983, Houston Texas), 179-193.

Appendix 1. Survey Questionnaire

USER REACTIONS TO EXISTING SYSTEMS

We would like to request your voluntary participation in this brief survey, the purpose of which is to test and refine a set of system rating scales. In the future these scales will be used to measure user reactions to new computer systems. For testing purposes, the survey asks about two existing systems, chosen simply because of their wide availability at the lab. Our interest is not in these systems directly, but rather in the statistical properties of the rating scales themselves. Your responses will remain completely anonymous.

Thank you for your participation.

Bill Groves
Rich Helms
Fred Davis (MIT)

How to use rating scales:

Today is a sunny day.

Strongly Disagree			Neutral			Strongly Agree
1	2	3	4	5	6	7

By circling the 6, you would be saying that you agree quite a lot with the given statement.

Sunny days are:

Neutral

Good |___|___|___|_X_|___|___|___| Bad

By placing an X in the center position on the scale, you would be saying that the given statement is neither good nor bad.

ELECTRONIC MAIL

Questions in this section concern your reactions to the use of electronic mail. By electronic mail we mean any mail sent via the computer system -- profs notes, messages, memos, files, and so on.

Usage of Electronic Mail

1. Electronic mail is currently available for me to use, if I want to. Yes ___ No ___ Not Sure ___

2. On the average, I use electronic mail (pick most accurate answer):

Don't use at all	Use less than once each week	Use about once each week	Use several times each week	Use about once each day	Use several times each day
___	___	___	___	___	___

If you don't use electronic mail at all, please skip to page 6.

3. I normally spend about ___ hours each week directly using electronic mail.

4. I have been using electronic mail for (pick most accurate answer):

less than 1 month	between 1 and 3 months	between 3 and 6 months	between 6 months and a year	more than a year
___	___	___	___	___

	Strongly Agree			Neutral			Strongly Disagree
5. I use electronic mail because I have chosen to, not because I am required to for my job.	1	2	3	4	5	6	7
6. I am quite knowledgeable about how to use electronic mail.	1	2	3	4	5	6	7

Overall Evaluation of Electronic Mail

All things considered, my using electronic mail in my job is:
(place X mark on each of the five scales)

Neutral

1. Good |____|____|____|____|____|____|____| Bad

2. Wise |____|____|____|____|____|____|____| Foolish

3. Favourable |____|____|____|____|____|____|____| Unfavourable

4. Beneficial |____|____|____|____|____|____|____| Harmful

5. Positive |____|____|____|____|____|____|____| Negative

Perceived Ease of Use of Electronic Mail

	Strongly Agree			Neutral			Strongly Disagree
1. I find the electronic mail system cumbersome to use.	1	2	3	4	5	6	7
2. Learning to operate the electronic mail system is easy for me.	1	2	3	4	5	6	7
3. Interacting with the electronic mail system is often frustrating.	1	2	3	4	5	6	7
4. I find it easy to get the electronic mail system to do what I want it to do.	1	2	3	4	5	6	7
5. The electronic mail system is rigid and inflexible to interact with.	1	2	3	4	5	6	7
6. It is easy for me to remember how to perform tasks using the electronic mail system.	1	2	3	4	5	6	7
7. Interacting with the electronic mail system requires a lot of mental effort.	1	2	3	4	5	6	7
8. My interaction with the electronic mail system is clear and understandable.	1	2	3	4	5	6	7
9. I find it takes a lot of effort to become skillful at using electronic mail.	1	2	3	4	5	6	7
10. Overall, I find the electronic mail system easy to use.	1	2	3	4	5	6	7

Perceived Usefulness of Electronic Mail

	Strongly Agree			Neutral			Strongly Disagree
1. Using electronic mail improves the quality of the work I do.	1	2	3	4	5	6	7
2. Using electronic mail gives me greater control over my work.	1	2	3	4	5	6	7
3. Electronic mail enables me to accomplish tasks more quickly.	1	2	3	4	5	6	7
4. Electronic mail supports critical aspects of my job.	1	2	3	4	5	6	7
5. Using electronic mail increases my productivity.	1	2	3	4	5	6	7
6. Using electronic mail improves my job performance.	1	2	3	4	5	6	7
7. Using electronic mail allows me to accomplish more work than would otherwise be possible.	1	2	3	4	5	6	7
8. Using electronic mail enhances my effectiveness on the job.	1	2	3	4	5	6	7
9. Using electronic mail makes it easier to do my job.	1	2	3	4	5	6	7
10. Overall, I find the electronic mail system useful in my job.	1	2	3	4	5	6	7

XEDIT

Questions in this section concern your reactions to the use of XEDIT.

Usage of XEDIT

1. XEDIT is currently available for me to use, if I want to. Yes ___ No ___ Not Sure ___

2. On the average, I use XEDIT (pick most accurate answer):

Don't use at all	Use less than once each week	Use about once each week	Use several times each week	Use about once each day	Use several times each day
___	___	___	___	___	___

If you don't use XEDIT at all, please turn to page 10.

3. I normally spend about ___ hours each week directly using XEDIT

4. I have been using XEDIT for (pick most accurate answer):

less than 1 month	between 1 and 3 months	between 3 and 6 months	between 6 months and a year	more than a year
___	___	___	___	___

	Strongly Agree			Neutral			Strongly Disagree
5. I use XEDIT because I have chosen to, not because I am required to for my job.	1	2	3	4	5	6	7
6. I am quite knowledgeable about how to use XEDIT.	1	2	3	4	5	6	7

Overall Evaluation of XEDIT

All things considered, my using XEDIT in my job is:
(place X mark on each of the five scales)

Neutral

1. Good |___|___|___|___|___|___|___| Bad
2. Wise |___|___|___|___|___|___|___| Foolish
3. Favourable |___|___|___|___|___|___|___| Unfavourable
4. Beneficial |___|___|___|___|___|___|___| Harmful
5. Positive |___|___|___|___|___|___|___| Negative

Perceived Ease of Use of XEDIT

	Strongly Agree			Neutral			Strongly Disagree
1. I find XEDIT cumbersome to use.	1	2	3	4	5	6	7
2. Learning to operate XEDIT is easy for me.	1	2	3	4	5	6	7
3. Interacting with XEDIT is often frustrating.	1	2	3	4	5	6	7
4. I find it easy to get XEDIT to do what I want it to do.	1	2	3	4	5	6	7
5. XEDIT is rigid and inflexible to interact with.	1	2	3	4	5	6	7
6. It is easy for me to remember how to perform tasks using XEDIT.	1	2	3	4	5	6	7
7. Interacting with XEDIT requires a lot of mental effort.	1	2	3	4	5	6	7
8. My interaction with XEDIT is clear and understandable.	1	2	3	4	5	6	7
9. I find it takes a lot of effort to become skillful at using XEDIT.	1	2	3	4	5	6	7
10. Overall, I find XEDIT easy to use.	1	2	3	4	5	6	7

Perceived Usefulness of XEDIT

	Strongly Agree			Neutral			Strongly Disagree
1. Using XEDIT improves the quality of the work I do.	1	2	3	4	5	6	7
2. Using XEDIT gives me greater control over my work.	1	2	3	4	5	6	7
3. XEDIT enables me to accomplish tasks more quickly.	1	2	3	4	5	6	7
4. XEDIT supports critical aspects of my job.	1	2	3	4	5	6	7
5. Using XEDIT increases my productivity.	1	2	3	4	5	6	7
6. Using XEDIT improves my job performance.	1	2	3	4	5	6	7
7. Using XEDIT allows me to accomplish more work than would otherwise be possible.	1	2	3	4	5	6	7
8. Using XEDIT enhances my effectiveness on the job.	1	2	3	4	5	6	7
9. Using XEDIT makes it easier to do my job.	1	2	3	4	5	6	7
10. Overall, I find XEDIT useful in my job.	1	2	3	4	5	6	7

My current job classification is:

_____ Management

_____ Secretarial

_____ Professional

My current job title is:

____

Feedback comments about this survey:

____

____

____

____

Thank you very much for your participation.

Appendix 2. Sign-up Sheet

BUSINESS GRAPHICS STUDY
SIGN-UP SHEET

NAME ______________________________

PHONE # ______________________________

Student ID # ______________________________

Please indicate your preferences for times and dates for participation in the hands-on demonstrations of the two business graphics systems. Beside each of the date/time slots listed below, indicate your preferences using the following codes:

i - impossible to come at that time
p- possible, but not convenient to come at that time
c- convenient to come at that time
v- very desireable/convenient time to come

Sat. 7/20 10am - 12 noon [ ]
Sat. 7/20 1-3pm [ ]
Sat. 7/20 4-6pm [ ]
Sat. 7/20 7-9pm [ ]

Sun 7/21 10am-12 noon [ ]
Sun. 7/21 1-3pm [ ]
Sun 7/21 4-6pm [ ]
Sun. 7/21 7-9pm [ ]

Mon. 7/22 10am-12 noon [ ]
Mon. 7/22 1-3pm [ ]
Mon. 7/22 4-6pm [ ]
Mon. 7/22 7-9pm [ ]

Tue 7/23 10am-12 noon [ ]
Tue 7/23 1-3pm [ ]
Tue 7/23 4-6pm [ ]
Tue 7/23 7-9pm [ ]

Wed 7/24 10am-12 noon [ ]
Wed 7/24 1-3pm [ ]
Wed 7/24 4-6pm [ ]
Wed 7/24 7-9pm [ ]

Thu 7/25 10am-12 noon [ ]
Thu 7/25 1-3pm [ ]
Thu 7/25 4-6pm [ ]
Thu 7/25 7-9pm [ ]

Fri 7/26 10am-12 noon [ ]

Appendix 3. Consent Form

Objectives of the Study

In this research study we would like to obtain your evaluations of 2 computer graphics products which are currently available on the market. The overall goal of the study is to develop techniques for measuring how well software systems fit the needs of potential users. As part of this effort, we are interested in comparing the effectiveness of two different techniques for demonstrating computer systems to potential users: videotape presentation and hands-on interaction.

Methods of the Study

The study involves two different sessions. In the first session, two different graphics systems will be presented to you by videotape. After viewing each system you will be asked to complete a questionnaire asking about your perceptions and attitudes toward the system. In the second session (which will be arranged at the time of the first session) you will receive a hands-on demonstration of the same two software systems. Again you will be asked to fill out a questionnaire regarding the two systems. After completing the second questionnaire, you will be given a $25.00 fee for your participation. Your individual responses will remain completely anonymous and will be aggregated by computer. No deception will be involved in the experiment. We are evaluating the graphics software itself and not you, the participants.

Disclaimer Regarding Videotapes

The videotapes to be used in the study were prepared by International Business Machines, Inc. (IBM). IBM was not involved in the development of the software products covered by the research study, and does not own or market them. The opinions expressed in the videotapes are IBM's and not those of the product owners. The opinions expressed were made solely for the purposes of this research study. The opinions may not be accurate, as they are intended to highlight an evaluation technique, and not to pass judgement on the value of the products. Accordingly, the product owners have not been given the opportunity to review and comment on the opinions. By signing this form you agree to keep the contents of the videotapes confidential.

I agree to participate in the research study as described on the previous page. I am free to discontinue participation in the study at any time, thereby waiving the $25.00 subject fee. If there are any questions I do not wish to answer, I may omit these. I agree not to discuss the contents of the experiment with any of my classmates until all of them have completed the entire procedure.

____ __________

Signature Date

Consent Form

Objectives of the Study

In this research study we would like to obtain your evaluations of 2 computer graphics products which are currently available on the market. The overall goal of the study is to develop techniques for measuring how well software systems fit the needs of potential users. As part of this effort, we are interested in comparing the effectiveness of two different techniques for demonstrating computer systems to potential users: videotape presentation and hands-on interaction.

Methods of the Study

You will receive a hands-on demonstration of two graphics software systems. After each system demonstration you will be asked to fill out a questionnaire regarding the system. After completing the questionnaires, you will be given a $25.00 fee for your participation. Your individual responses will remain completely anonymous and will be aggregated by computer. No deception will be involved in the experiment. We are evaluating the graphics software itself and not you, the participants.

I agree to participate in the research study as described above. I am free to discontinue participation in the study at any time, thereby waiving the $25.00 subject fee. If there are any questions I do not wish to answer, I may omit these. I agree not to discuss the contents of the experiment with any of my classmates until all of them have completed the entire procedure.

Signature ______________________ Date ________

CHART-MASTER Instruction Booklet

1. CHART-MASTER Overview

CHART-MASTER$_{tm}$ is a graphics software package for the IBM Personal Computer (PC) designed by Decision Resources, Inc. of Westport, Connecticut. CHART-MASTER is intended for displaying numerical information in graphs such as bar charts, line charts and pie charts. The sample booklet gives examples of various types of graphs made with CHART-MASTER.

2. Learning to Use CHART-MASTER

To introduce you to CHART-MASTER, we will walk you through the process of creating a fairly simple chart, and later explain some of the details involved with more complicated charts. The chart we will construct looks like the one shown in figure 1.

We ask that you follow the instructions in a step-by step fashion to make sure you get introduced to all of the features discussed. The instructions should be self-explanatory. However, the experimenter will be present to answer any questions if necessary. You may move through the material in this instruction manual at your aow pace, but **please follow the instruction closely**.

Personal Computer Keyboard

There are a few things you will need to know about the IBM PC keyboard in order to follow this instruction book. First, you will be using the "**enter**" key a lot. This is the key that has a bent arrow pointing to the left (   ) located to the right of the letter keys. The **enter** key is used for sending your responses from the keyboard to the personal computer's processing unit.

Second, you will need to use the "**shift**" keys to type upper case letters. These are the keys located on both sides of the letter keys that each have a thick arrow pointing upward (   ). You can lock the shift key into upper case using the "**Caps lock**" key. Third, the "**backspace**" key, located above and right of the letter keys and having a left pointing arrow (   ), is used to retype text that has been incorrectly entered. Fourth, there are a set of **arrow keys** located on the right of the keyboard. These will be used to specify type sizes and styles for text on your chart, and will be explained later when we introduce text options.

Main Menu

When you first start out with CHART-MASTER, you will see the main menu. Since we are going to make a new chart, select menu option 1: "**Create a Chart**". Do this by typing a "**1**" and hitting the "**enter**" key.

Chart Title

In creating a chart, CHART-MASTER will promt you for the title of your chart (see figure 2). For this example, specify the first line of the title to be: XYZ CORP; type **XYZ CORP**, and hit [**enter**]. Similarly, specify the second line to be: "**PRODUCT SALES**", and the third line: "**JUNE-SEPTEMBER 1984**". Leave the fourth title line blank by simply hitting [**enter**] without typing anything. If you make a mistake typing, you can simply [**backspace**] over the wrong letters and retype them. If you already entered the incorrect data, don't worry- there is a way of changing it after you enter the rest of the chart information.

Axis Labels

CHART-MASTER will next ask you to give the labels for the horizontal (x) and vertical (y) axes of your chart (see figure 2). Each axis label can be up to two lines, which are specified in much the same way as the chart title was. For our example, enter the following labels (remember to hit [enter] after each response):

ENTER X-AXIS LABEL: **Month**

ENTER SECOND X-AXIS LABEL: **1984**

ENTER Y-AXIS LABEL: **THOUSANDS OF DOLLARS**

ENTER SECOND Y-AXIS LABEL: Hit [**enter**]

Data Entry Mode

CHART-MASTER will ask you how you would like to enter your data, either manually (M) or via the Document Interchange Format (D). For the purposes of this hands-on demonstration, we will only use manual input. Therefore, type **M** and hit **[enter]**.

Number of Variables

In CHART-MASTER, variables refer to the categories being measured, such as products or regions. In our example, we will be comparing two products (see figure 2). Therefore, you should specify "**2**" variables, and hit **[enter]**.

Number of Observations

Each point along the x-axis is called an "observation". These are often time periods, such as weeks, months, or years, but can be other groupings also, such as departments, regions, etc. In our case, we want to graph sales for our two products over the 4 months: June, July, August and September (see figure 2). Therefore, specify that there are "**4**" observations, and hit **[enter]**.

Variable Labels

Specify "Product 1" and "Product 2" as the labels for the two variables, hitting **[enter]** after each. These will appear later in the chart legend (figure 2).

Observation Labels

Specify the first observation label to be: 1) **JUNE**. If you type it in all capital letters and repeatedly hit **[enter]**, CHART-MASTER will automatically supply the rest of the months- **JULY, AUGUST,** and **SEPTEMBER**.

Data Entry

The next step is to enter the data values of each observation (month) for each of the two variables (products). CHART-MASTER will first ask you for the 4 observations for Product 1. Type in **38, 33, 35,** & **31** for the 4 observations. For Product 2, type in **18, 28, 37** and **46**.

Chart Verification

After entering all the chart data, CHART-MASTER will return to the main menu. You may now verify the data you have just entered by selecting (2) "**Verify Chart**". Specify "S" to see the chart information on the screen. Press **[enter]** to "scroll"through the information. Repeatedly hitting "enter" will eventually get you back to the main menu.

Chart Editing

If you made a mistake typing in the chart informaton, you can correct it with the Edit function. You can edit the titles, legends, labels or data. Select **(4) Edit** from the main menu. You now see the **Edit Menu** on the screen. This menu allows you to select which category of information in the chart specification you wish to modify. After selecting the desired category, a sub -menu may appear asking you to specify further which characteristic to change. Finally, it will give you the old information and ask you to type in the new information.

You should now edit any mistakes in your example chart information. The Edit function is also helpful on occasions where you would like to update old information, or, as we will see later, for adding new variables or observations and for changing the font and size of titles and labels.

Chart Plotting

In order to see your chart on the screen, you must "plot" it. Do this by selecting the **Plot Chart** command **(3)** from the main menu. This will cause the "Plot Menu" to be displayed, showing the six different ways to plot your chart. Try plotting your data as a **Clustered Bar Chart** by selecting **(1)**.

The next menu asks you to specify which option to use when plotting your chart. The first three options will plot the chart on the screen, either in high resolution black and white (1), or medium resolution using either of two different color schemes (2 & 3). The second three plot the chart on various types of output devices: (4) Polaroid Pallette (for making 35mm slides), (5) a plotter (for high quality hardcopy or transparencies) or (6) a printer (for medium quality hardcopy or transparencies). For this experiment, we do not have any hardcopy output devices attached to the PC's, so you will have to plot the charts on the screen only. The sample charts you have seen were made using a printer.

Specify **(2) Medium Resolution: Green/Red/Brown** to view your chart on the screen. You should see your chart on the screen.

Now try to plot you chart as a stacked bar chart. In order to do this, you will want to return to the plot menu. First press the escape key **[esc]** to return to the main menu, and select **(3) Plot Chart**. This time, specify **(2) Stacked Bar Chart**, and to see another color scheme, select **(3) Medium Resolution: Cyan/Magenta/White**.

In a similar manner, try plotting your chart as a scatter chart, a pie chart, and an area chart. Remember, to get back to the main menu from a chart screen, press the **[enter]** button.

Adding a New Variable

Now we would like for you to add a new variable called "Product 3". This is done via the Edit Menu. From the Main Menu, select **(4) EDIT** to get into the edit menu. Then select menu item **(6) EDIT ALL DATA FOR ONE VARIABLE**. On the next menu, select option **(1) ADD A VARIABLE**. The next menu will show your existing variable labels, and asks you **"WHICH VARIABLE"**. Specifying (3) will add the new variable to the list. You will be asked by CHART-MASTER to specify the name of the new product.

Next, you will be asked for the Product 3 observations, which are **15, 21, 25** and **18**. Try plotting a clustered bar chart, a stacked bar chart, a line chart, and a pie chart reflecting these changes.

Changing Text Options

Assume we would like to make the title of our example chart a larger size and change the type style (font). This requires that we enter the **Edit Menu** (select 4 from Main Menu). Select **(1) TITLES**. The 3 lines of the Title (and fourth blank line) will be displayed. Specify **(1)** to change the first line. At the bottom of the screen you will notice 6 options (font, size, italics, justify, pen and underline). To change the values of these format options, you need to position the "cursor" to the desired option. This is accomplished by hitting the **[PGDN]** key located at the right of the keyboard. This will cause a set of brackets to appear around the notation "**STD**" under fonts. This means that fonts is the current option. Use the [**up arrow**] key to "scroll" through the font options. Select bold roman by leaving the option at (**BRM**) when it. Now, to move the brackets over to the size option, hit the [**right arrow**] key. Again using [**up arrow**] select size **10**. We are now done changing the options for the first line of the title, so hit [**PGUP**] to return to the top of the screen. You must now **retype line 1** of the title to complete the process.

Using a similar approach, try changing the second line to **bold** font **(BOL)** in **italics**. This is the way to customize any of the titles or labels in a chart. Plot the chart now to see the effect of these changes.

Chart Storage and Retrieval

It is often desireable to store the chart information so that if later you want to plot the chart again, perhaps using a different format, you will not have to re-enter all the information. To try storing your chart information. On the Main Menu, select **(5) STORE/RETRIEVE/DELETE CHART**. On the Storage Menu, select **(1) STORE CHART**. Next, enter the name you would like to give the chart (why not **SALES**?). and hit **[enter]**. When asked to insert disk C, just ignore it and hit **[enter]**.

To verify that your chart has been properly stored, you may select **(4) Chart Catalog**. You should see your chart listed.

Specifying Options

Using CHART-MASTER you can custom-tailor any chart by changing the "Options" governing its format. Each chart has automatic or "default" format characteristics. However, you can change these using the Options Menu. Any changes you make will remain in effect until you change them again. The

Options menu consists of seven "pages" (screens). Each page of the options menu corresponds to a different type of chart, as shown below:

Page 1	All Charts
Page 2	Bar Charts
Page 3	Line Charts
Page 4	Scatter Charts
Page 5	Left Y-axis
Page 6	Right Y-axis
Page 7	Pie Charts

To give you a flavor for how Options are specified, we will go through a few examples.

Bar Chart Options

Consider the example chart data we have been working with. Assume that we would like to make two changes to how a clustered barchart is plotted using this data. First, we would like to have the data values corresponding to each bar printed atop the bar. Second, we would like to have the bars plotted horizontally instead of vertically.

To make these modifications, first select **(6) CHANGE OPTIONS** from the main menu. You will see the first of the seven Option Menu pages (the page number appears near the top of the screen). Since the changes you want to make involve the bar charts, you need to move to page 2. This is done by hitting **[enter]**. If you look at the first item on page 2 of the Options Menu: **"PRINT DATA ATOP BARS?"**, you will see that its present setting is **"NO"**. This is the default setting. To change it to **"YES"**, just type 1 and hit **[enter]**. Like many of the options settings, this feature is like a toggle switch- selecting it will change if back and forth between its possible values. Similarly, select **(6) HORIZONTAL BARS** to plot the bars horizontally.

Now plot the barchart to see the effect of these changes. First, to get back to the main menu, select **(7) QUIT OPTIONS**. Now you can plot the chart (as you have done before) to observe the new chart format.

Pie Chart Options

Now we can try two changes on the pie chart format. Recall that when you plotted the sample data as a pie chart before the four pies corresponding to the four months were the same size. An alternative way to plot them is to let their size vary according to the relative size of the total sales for the month. This is called a proportional pie chart. To make pie charts proportional, first move to the **seventh page** of the **Options Menu**. Then, select **(8) PROPORTIONAL PIES** to change the setting from "**NO**" to "**YES**". Exploded slices are one or more of the slices of a pie chart that are "pulled out" of the pie to highlight them. Explode the pie slice corresponding to the Product 3 by selecting **(7) EXPLODED SLICES.** You will then see another menu to specify whether you want None, All or Selected pie slices to be exploded. Choose **(3) Selected**. Then for each of the products you will be asked whether or not to explode its slice. Answer yes for **Product 3**.

Now plot a pie chart using the sample data to see the changes you have made via the Options Menu.

Other Options

Although there is not enough time to try out all of the available options during this hands-on demonstration. we do have enough time to briefly explain some of them. You can change the size, orientation (horizontal or vertical) and location of you chart on a screen or hardcopy page. You can omit the frame that surrounds the charts.

You can prevent CHART-MASTER from "cross-hatching" the bar or pie charts, or specify the crosshatch patterns for particular bars or pie slices. You can change the location and format of the Legend that explains the variables.

For line charts, you can have the individual data points depicted in any of a variety of symbols. You can fit linear regression and other kinds of trend lines to your charts.

The scales on the Y-axis can be changed from linear to logarithmic. A separate y-axis on the right side of a graph can be chosen, with selected variables plotted against it while the other variables are plotted against the left y-axis.

In summary, there is quite a lot of variety in chart format that can be attained with the Options Menu of Chart-master.

This concludes the self-paced instruction process. Feel free to experiment with using CHART-MASTER until the experimenter asks you to stop and fill out a questionnaire.

PENDRAW Instruction Booklet

1. PENDRAW Overview

PENDRAW$_{tm}$ is a graphics software package designed by Pencept, Inc. of Waltham Massachusetts to be used with the IBM$_{tm}$ Personal Computer (PC). Primarily intended as a tool for making visual aids for business presentations, PENDRAW enables the user to create a wide variety of diagrams, charts, drawings and geometrical shapes. The sample book gives examples of various types of visual aids made with PENDRAW.

2. Learning to Use PENDRAW

To introduce you to PENDRAW, we will walk you through several of its functions and capabilities. After that, you will be guided through the steps of making an example graph using PENDRAW.

We ask that you follow the instructions in a step-by-step fashion to make sure you get introduced to all of the features discussed. You may move through the material in this instruction manual at your own pace, but **please follow the instructions closely.**

Personal Computer with Digitizer Tablet

The PENDRAW software runs on an IBM PC using a digitizer tablet. The tablet substitutes for the keyboard normally used with the PC. The tablet has two regions, the command **template**, which is the plastic card located near the top of the tablet, and the graph paper **form**, located below the template. The user tells PENDRAW what to do next by selecting commands on the template using the electronic **pen**. When the user writes on the form, the tablet captures his or her penstrokes and transfers them onto the screen. If at any point during the experiment you would like a clean form, just ask the experimenter.

Calibration

First the PENDRAW title screen will ask you to calibrate the PENDRAW template, which is the plastic rectangle near the top of the tablet. Calibration tells the program exactly where the template is positioned on the tablet.

Follow the instructions on the bottom of the screen to calibrate your template. When you are asked to touch the corners of the template, be sure your pen tip touches the center of the small circles.

FreehandDrawing

After calibration, PENDRAW automatically goes into **Freehand** mode. The current mode is always displayed at the bottom of the screen. Try writing on the paper form to see how freehand drawing works.

As you move your pen about 1/2 inch above the surface of the paper you will notice a " + " moving on the screen. This is called the "cursor" and serves to show you the location of the pen before you touch the paper.

Lines

Now select **line mode** by touching your pen down on the template box labeled **line** (in the **shapes** block). Notice that the bottom of your screen now indicates you're in line mode. Try drawing some lines. If the screen gets too messy, you can clear the screen at any time by selecting **clear** on the template.

PENDRAW will make a line between the point where you set the pen down onto the tablet to the point where you lift the pen off again.

Boxes

Select **Box** and experiment making boxes. The point at which you initially touch your pen to the paper anchors a corner of the box; the direction you move from that point determines the shape of the box; the distance you move from that initial point determines the size of the box. **Clear** your screen again if you wish.

Circles

Now select **Circles**. Touch your pen to the paper to define the center of the circle and move the pen diagonally to define the shape and size. Experiment making circles. Notice that the more perfectly diagonal you make your penstroke, the more perfectly round your circle will be.

Changing Color

Notice the green color bar on the bottom of the screen. Select **Color** on the template. Four color alternatives will appear at the top of the screen. Move the cursor to the color **red** by holding the pen above the surface of the paper. To select the color red, touch the pen to the paper while the pen is in the red region. Notice that the color bar at the bottom of the screen has changed to red. Try drawing red circles now.

Changing Linewidth

Select **Linewidth**. Notice that a rectangular box is displayed at the top of the screen. The bottom half of the rectangle contains samples of the seven linewidths you can choose. **Move the cursor** to the line thickness you would like to select and touch your pen to the paper. The selected linewidth is now displayed in the upper left section of the rectangle. To confirm your choice **click the pen button**. Use your new linewidth to draw circles.

Notice that as you can select any of the available shapes: freehand; line; circle, or box, the color and linewidth you have selected remains the same until you change them. You may want to experiment with this a little.

Undo

Selecting the **undo** function will clear anything off the screen that you put there since either the last time you changed commands (e.g. circle to box) or the last time you clicked the button on the pen. When you are creating a drawing, it is sometimes handy to click the button on the pen to save a particular screen, so that in case you make a mistake while adding additional details, you can select the undo function and return to the "saved" screeen . To familiarize yourself with this, add a circle to the screen, click the button, add another circle, and then select undo. Add another circle, select box, add a box, and select undo again.

Filling a Region

The **Fill** function enables you to "color in" any enclosed region with a color of your choice. To see how this works, draw a circle, select the **Fill** function, position the pen a half inch above the circle (by locating the cursor within the circle), and touch the pen down on the paper. You can change colors for the **Fill** function by selecting **Colors** on the template, as before.

Erase

Erase allows you to erase specific areas of the screen without erasing the entire screen. When you select Erase, at the bottom of the screen PENDRAW will ask you to "**define box**". You specify which area you would like to erase by enclosing it with a box. The erase box is specified in much the same way as done with the **Box** function. That is, the place where the pen is first touched down on the paper anchors one corner of the box, and the point where the pen is lifted off the paper again defines the point of the box diagonally across from the anchor point. The defined region is erased as soon as you lift up the pen. Try using **Erase** on the circle defined in the last paragraph. If you make a mistake using **Erase**, you can recover the original screen using the **Undo** function.

When using erase, you may specify a **color** option. Whatever color you select is the color that is left in the area that you erased information from. The usual color used for erasing is black. Other colors are helpful when you are erasing something from a colored region.

Move

Move allows you to move a specified region of a screen to another location on the screen. With **Move**, you define the region to be moved in the same way as you defined regions using **Erase**: by drawing a box around it. After selecting the **Move** function, PENDRAW will ask you to define the box you would like to move. Next, it will ask you to point to the location you would like to move it to. You do this by touching the pen down on the paper. You may continue moving the region around until you have it in the desired place, and then **click the pen button** to freeze it into place. Experiment with **Move** by drawing a couple of circles or boxes and moving them around on the screen.

Copy

The **Copy** function works on much the same principle as **Move**, with the exception that it will leave the original copy of the defined area intact, making a new copy of it in the designated location. Otherwise, you define the box, point to the target location, and freeze the copy into place in the the same way as done with the Move function. Try using the Copy function on the some of the circles and boxes on your screen. Remember, you can use the **Undo** function if you are not satisfied with your results, and the **Clear** function if your screen gets too messy.

Text

This function allows you to add text to your chart. Select the **Text** function on the template. The **Text** function works by recognizing characters that you write into boxes on the graph paper form. Therefore you should calibrate the form to let the software know its exact position. Do this by selecting the **Form** option under **Calibrate** in the upper right hand corner of the template. Next, touch the pen on the corners of the form as instructed at the bottom of the display screen.

If you now hold the pen a half inch above the paper, you should find that the " + " sign cursor is missing. That's because in **Text** mode a different cursor approach is used. If you **hold in the pen button**, you should see a nall block of light appear on the screen. This is used to indicate where you would like your text to be positioned. Keeping the pen button pressed, you can move this block of light around on the screen. **Release the button** to freeze it in place.

Now, try writing an "**A**" in any grid box on the form, it should appear on the screen in place of the block of light. In order for PENDRAW to recognize your hand printed characters, you will need to print them in capital, block-letter style. Table 1 gives a guide about how to form your characters so that PENDRAW will understand them. It usually requires a little practice before you are able to write characters so that PENDRAW will understand a high percentage of them. If PENDRAW does not properly recognize a letter, you can try again, simply writing over the first letter. If you want to erase a letter altogether, just write an upside down "**U**" over the letter.

If you were successful in writing an "**A**", try writing a "**B**" next to it, by printing a "**B**" in the next grid box to the right of the "**A**". You may want to spend a few moments experimenting with **Text** at this point

At the bottom of ther screen you should see **Text: A-Z**. This means that you are restricted to upper case letters. In the **Annotate** block of the template there is a cluster of 4 function boxes surrounding the label **Character Set**: **Numeric, All, Upper Case** and **Lower Case**. These functions are used to define the characters to be used. Selecting **All** enables you to write both letters and numbers. If you wish to write lower case text, select **Lower Case**, but continue to print the input text on the tablet using capital block letters.

Title

Title is similar to **Text** except that it allows you a variety of type fonts (styles) and sizes. Like **Text, Title** works by recognizing your hand-printed letters. Unlike **Text,** however, **Title** will accept letters of varying sizes, using the size of your hand-printed letter to determine the size of the **Title** letters. Therefore, you don't have to print your letters inside grid boxes when using **Title**.

Select the **Title** function on the template. To choose a type style (font) by selecting **Fonts** on the template. A list of available fonts will appear on the screen (these appear automatically the first time you use **Titles**). To choose one, hold the pen button down while moving the pen down the list above the surface of the tablet to "**drag**" the arrow down the list. When it points to **Sanserif**, release the button to freeze the arrow. **Press and release the pen**

button to see a sample of the selected font. **Click the pen button again** to confirm your choice.

With **Title**, you need to pause between each letter to give the software time to translate your letter into the desired font. Try printing an "**A**" about two inches high on the form. As with **Text**, you can change an incorrect letter by writing over it, or erase it by writing an upside down "**U**" over it. Print the "**BCD**" after the "**A**" in the same manner.

Both **Text** and **Title** can be used in conjunction with several of the other functions. For example, you can change the **Color** option and **Move** or **Copy** text or titles. Try these if you like.

Sample Exercise: Creating a Graph

Now we would like you to try creating a graph. Imagine that you are planning a management presentation showing your company's quarterly sales figures for 1984. The sales figures to be graphed are as follows:

first quarter	$500,000
second quarter	250,000
third quarter	600,000
fourth quarter	900,000

We would like to walk you through the process of constructing a barchart of these sales figures. There will be four bars representing the four quarters, with their heights representing the sales amounts for each quarter (as shown in figure 1).

Axes

First, **Clear** the screen. Next create the horizontal and vertical axes of the chart. To do this, draw horizontal and vertical **Lines** as shown in figure 2. Once you have them the way you want them, click the pen button to save them.

Next, mark off the money increments on the vertical axis. Since we know that the highest number to be charted is just under a million, let's make the top hashmark the million dollar point. Just make a small horizontal **Line** going off to the left of the vertical axis as shown on figure 3. Similarly, draw the 1/4, 1/2 and 3/4 hashmarks.

Bars

Next, draw the bars themselves, using the **Box** feature. Position the first box so that it's bottom is aligned with the horizontal axis, and it's top is even with the second hashmark, the half-million point. Once you position the bar the way you want it, click the pen button to save it. Similarly, draw the three other bars so that their heights correspond to the correct sales figures on the vertical axis, as shown on figure 4. Now, color the bars red by: selecting the **Fill** function, selecting the **Color** function, specifying **red**, and then **touching the inside of each bar** with the pen.

Labels

Now we would like you to put labels on the graph. You are going to put Q1, Q2, Q3, and Q4 under the four bars of the graph to designate the four quarters, as shown in figure 5. Select the **Text** function on the template. Select **All** in the **Character Set** box so that you will be able to write both letters and numbers. Next, keeping the pen button pressed, move the block of light under the first bar on the chart and release the button to freeze it in place.

You are now ready to input the text. If you write a "**Q**" in any grid box on the form, it should appear in place of the block of light. After successfully writing the "**Q**", write a "**1**" in the grid box next to it. If you are satisfied with "**Q1**", click the mouse button to save it (in case you need to use **Undo** later). Similarly, reposition the block of light under the next bar and write "**Q2**" and so on until all of the bars have been labeled.

Now put labels on the vertical axis of the chart: $1M, .75, .5 and .25 respectively (see figure 5). You will probably want to click the pen button after successfully creating each of these in order to save them.

Title

Finally, you will want to put a title on the new chart - SALES 1984, as shown in figure 1. For this, select the **Title** function. Choose the **Romanbld** font by dragging the arrow down the page. Print an "**S**" about an inch high above your chart. As with **Text**, you can change an incorrect letter by writing over it or erase it by writing an upside down "**U**" over it. Print the "**ALES**" after the "**S**" in the same manner. If you like, you can adjust the position of the title using the **Move** function. Use **Title** again to add the year **1984** to your title.

Normally, you would be able to print out your new chart (selecting the **Print** function). However, no printer is hooked up to these test Personal Computers. However, the sample book gives examples of what it might look like if you made a paper printout of your graph.

Saving the Chart

Now that you're done creating your chart, you should save it for future reference. To do this, select the **Save** function. A graphical screen such as the one you have just created is referred to as a "slide". You will need to specify a name for the new slide in order to store it. Why not call it **SALES**? You just need to **print the name** in any set of consecutive grid boxes on the grid form. Once the correct name is properly entered, click the pen button. The software will tell you that the file will be stored in "compressed format" unless you change the option. Select **Save** again to complete the process of storing the chart.

This concludes the self-paced instruction process. Feel free to experiment with using PENDRAW until the experimenter asks you to stop and fill out a questionnaire.

Appendix 5. **Experiment Questionnaire**

Student ID # ________________________________

Instructions:

When filling out this questionnaire, answer the questions from the perspective of either your **current job** (if full-time employed), your **last full-time job** before becoming a full-time student, or the **job you expect to have** after going back to work full-time.

How to use rating scales:

Tomorrow will be a sunny day.

likely |_____|_____|_____|_____|_____|_____|_____| **unlikely**

extremely quite slightly neither slightly quite extremely

By placing an X mark as shown, you would be saying that you believe it is quite unlikely to be sunny tomorrow.

For tomorrow to be sunny would be:

Good |_____|_____|_____|_____|_____|_____|_____| **Bad**

extremely quite slightly neither slightly quite extremely

By placing an X mark as shown, you would be saying that, for you, a sunny day tomorrow would be extremely good .

Important, please:

Read each question carefully before answering.

Give an answer for every scale.

A. Overall Evaluation

My using CHART-MASTER in my job would be:
(place an X mark on **all four** scales)

Good	extremely	quite	slightly	neither	slightly	quite	extremely	**Bad**
Harmful	extremely	quite	slightly	neither	slightly	quite	extremely	**Beneficial**
Wise	extremely	quite	slightly	neither	slightly	quite	extremely	**Foolish**
Negative	extremely	quite	slightly	neither	slightly	quite	extremely	**Positive**

How confident are you in the ratings that you have made on this page?

Not at all Confident |_______|_______|_______|_______|_______|_______|_______| **Completely Confident**

B. Perceived Ease of Use

Learning to operate CHART-MASTER would be easy for me.

likely	extremely	quite	slightly	neither	slightly	quite	extremely	unlikely

I would find it easy to get CHART-MASTER to do what I want it to do.

likely	extremely	quite	slightly	neither	slightly	quite	extremely	unlikely

My interaction with CHART-MASTER would be clear and understandable.

likely	extremely	quite	slightly	neither	slightly	quite	extremely	unlikely

I would find CHART-MASTER to be flexible to interact with.

likely	extremely	quite	slightly	neither	slightly	quite	extremely	unlikely

It would be easy for me to become skillful at using CHART-MASTER.

likely	extremely	quite	slightly	neither	slightly	quite	extremely	unlikely

I would find CHART-MASTER easy to use.

likely	extremely	quite	slightly	neither	slightly	quite	extremely	unlikely

How confident are you in the ratings that you have made on this page?

Not at all Confident |_______|_______|_______|_______|_______|_______|_______| **Completely Confident**

C. Perceived Usefulness

Using CHART-MASTER in my job would enable me to accomplish tasks more quickly.

likely |____|____|____|____|____|____|____| **unlikely**
extremely quite slightly neither slightly quite extremely

Using CHART-MASTER would improve my job performance.

likely |____|____|____|____|____|____|____| **unlikely**
extremely quite slightly neither slightly quite extremely

Using CHART-MASTER in my job would increase my productivity.

likely |____|____|____|____|____|____|____| **unlikely**
extremely quite slightly neither slightly quite extremely

Using CHART-MASTER would enhance my effectiveness on the job.

likely |____|____|____|____|____|____|____| **unlikely**
extremely quite slightly neither slightly quite extremely

Using CHART-MASTER would make it easier to do my job.

likely |____|____|____|____|____|____|____| **unlikely**
extremely quite slightly neither slightly quite extremely

I would find CHART-MASTER useful in my job.

likely |____|____|____|____|____|____|____| **unlikely**
extremely quite slightly neither slightly quite extremely

How confident are you in the ratings that you have made on this page?

Not at all Confident |____|____|____|____|____|____|____| **Completely Confident**

D. Anticipated use of CHART-MASTER

For questions on this page, assume CHART-MASTER would be available for you to use on your current (or future) job.

Assuming CHART-MASTER would be available on my job, I predict that I will use it on a regular basis in the future.

likely	extremely	quite	slightly	neither	slightly	quite	extremely	**unlikely**

improbable	extremely	quite	slightly	neither	slightly	quite	extremely	**probable**

In my job, I am most likely to use CHART-MASTER (pick one):

[ ]	[ ]	[ ]	[ ]	[ ]	[ ]
not at all	**0-5 hours per week**	**5-10 hours per week**	**10-15 hours per week**	**15-20 hours per week**	**more than 20 hours per week**

What are the chances in 100 that you will become a CHART-MASTER user?

0	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**

How confident are you in the ratings that you have made on this page?

Not at all Confident |_______|_______|_______|_______|_______|_______|_______| **Completely Confident**

E. Perceived Characteristics of Output

Assuming I were to use CHART-MASTER, the quality of the output I would get would be high.

likely	extremely	quite	slightly	neither	slightly	quite	extremely	unlikely

Using CHART-MASTER, the effectiveness of the finished product would be:

low	extremely	quite	slightly	neither	slightly	quite	extremely	high

The charts and graphs I would make with CHART-MASTER would be professional-looking.

likely	extremely	quite	slightly	neither	slightly	quite	extremely	unlikely

How confident are you in the ratings that you have made on this page?

Not at all Confident								Completely Confident

F. Anticipated Enjoyment of Using CHART-MASTER

I would find using CHART-MASTER to be enjoyable

likely I_______I_______I_______I_______I_______I_______I_______I **unlikely**
extremely quite slightly neither slightly quite extremely

Using CHART-MASTER would be

pleasant I_______I_______I_______I_______I_______I_______I_______I **unpleasant**
extremely quite slightly neither slightly quite extremely

I would have fun using CHART-MASTER

likely I_______I_______I_______I_______I_______I_______I_______I **unlikely**
extremely quite slightly neither slightly quite extremely

How Confident are you in the ratings you have made on this page?

Not at all Confident I_______I_______I_______I_______I_______I_______I_______I **Completely Confident**

Numeric Charts

Numeric charts are charts or graphs that are used to present numerical information in a visual format, and include pie charts, bar charts, line charts, scatter charts.

Importance and Relevance

In my job, **numeric charts** are:

unimportant |_____|_____|_____|_____|_____|_____|_____| **important**

extremely quite slightly neither slightly quite extremely

relevant |_____|_____|_____|_____|_____|_____|_____| **irrelevant**

extremely quite slightly neither slightly quite extremely

CHART-MASTER

I would rate the **difficulty** of making **numeric charts** using **CHART-MASTER** as:

|_____|_____|_____|_____|_____|_____|_____|

impossible **moderate effort** **effortless**

I would rate the **quality** of **numeric charts** made with **CHART-MASTER** as:

|_____|_____|_____|_____|_____|_____|_____|

unacceptable (or non-existent) **moderate quality** **professional quality**

PENDRAW

I would rate the **difficulty** of making **numeric charts** using **PENDRAW** as:

|_____|_____|_____|_____|_____|_____|_____|

impossible **moderate effort** **effortless**

I would rate the **quality** of **numeric charts** made with **PENDRAW** as:

|_____|_____|_____|_____|_____|_____|_____|

unacceptable (or non-existent) **moderate quality** **professional quality**

Non-numeric Charts

Non-numeric charts are charts that do not involve the presentation of numeric information, but include diagrams, geometrical shapes etc. Examples of these include, organizational charts, flowcharts, and drawings.

In my job, **non-numeric charts** are:

unimportant |_____|_____|_____|_____|_____|_____|_____| **important**
extremely quite slightly neither slightly quite extremely

relevant |_____|_____|_____|_____|_____|_____|_____| **irrelevant**
extremely quite slightly neither slightly quite extremely

CHART-MASTER

I would rate the **difficulty** of making **non-numeric charts** using **CHART-MASTER** as:

|_____|_____|_____|_____|_____|_____|_____|
impossible **moderate effort** **effortless**

I would rate the **quality** of **non-numeric charts** made with **CHART-MASTER** as:

|_____|_____|_____|_____|_____|_____|_____|
unacceptable (or non-existent) **moderate quality** **professional quality**

PENDRAW

I would rate the **difficulty** of making **non-numeric charts** using **PENDRAW** as:

|_____|_____|_____|_____|_____|_____|_____|
impossible **moderate effort** **effortless**

I would rate the **quality** of **non-numeric charts** made with **PENDRAW** as:

|_____|_____|_____|_____|_____|_____|_____|
unacceptable (or non-existent) **moderate quality** **professional quality**